Developing methods to generate single-sex litters

Douglas, Charlotte; Turner, James

Awarding institution:
King's College London

Developing methods to generate single-sex litters

Charlotte Louise Douglas

King's College London

and

The Francis Crick Institute

PhD Supervisor: Dr James Turner

A thesis submitted for the degree of

Doctor of Philosophy

King's College London

September 2019

Declaration

I, Charlotte Douglas confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Producing single-sex litters can have a positive impact for many applications, including reducing disease-spreading vectors, increasing effective breeding populations for conservation, in agricultural biology, and in research; reducing unnecessary animal culling, in accordance with the Home Office 3Rs. In the laboratory mouse model, *Mus musculus,* males carry the sex chromosome complement XY. The daughters inherit the X chromosome, and the sons inherit the Y chromosome. In this thesis I exploit the unique inheritance of the X and Y chromosome from the male heterogametic sex to investigate multiple methods of generating all-male or all-female litters. Co-inheritance of two alleles, one from the father and one from the mother, which only upon the presence of both alleles causes embryonic lethality, was called a bi-component system. I first performed a proof-of-principle strategy utilising two pre-existing mouse models, X-Cre and an inducible *Rosa26*-DTA. Diphtheria toxin A (DTA) expression was only active upon excision of a floxed-STOP cassette by Cre recombinase. Co-inheritance of the X-Cre allele from the father, and *Rosa26*-DTA allele from the mother resulted in activation of the DTA and female-specific embryonic non-viability. In an alternative approach, I used CRISPR-Cas9 to generate mutations in a target housekeeping gene by co-inheritance of a Cas9 and sgRNA transgene. I first investigated the mutation efficiency of multiple sgRNAs *in vitro,* targeting an essential gene *Topoisomerase 1*. Highly-mutagenic sgRNA constructs were then carried forward to generate sgRNA-expressing mouse lines via recombination-mediated transgenesis. The sgRNA-expressing mice were mated with a pre-existing *Rosa26*-Cas9 mouse line. I showed that co-inheritance of the Cas9 and sgRNA transgene induced mutations in the target *Topoisomerase 1* locus. Co-inheritance of the two transgenes also resulted in embryonic non-viability due to loss-of-function of *Topoisomerase 1,* whilst inheritance of a single transgene did not result in loss-of-function mutations. Lastly, in order to generate sex-specific mutations by CRISPR-Cas9, I generated sex chromosome-linked Cas9 transgenes. I showed that the X-Cas9 transgene expressed Cas9 *in vitro* and *in vivo*, and in the future this line may be a useful tool for generating mutations in a target gene in a sex-specific manner.

Acknowledgements

This thesis would not have been possible without the input, contribution and support from many people, and for that I am extremely grateful.

First and foremost I would like to thank James Turner, without whom this project would not have existed, and who has provided an endless supply of great scientific advice and discussion, and who has truly been a great mentor. I am also extremely grateful to all Turner lab members, past and present; Andrew, Alex, Ben, Bryony, Daniel, Elias, Fanny, Jasmin, Mahesh, Oana, Sergio, Shantha, Sugako, Ruta, Takayuki and Valdone; for all of their wonderful scientific discussions, suggestions, and contributions to this thesis. In particular, I would just like to highlight a few people that have been involved more heavily in this project. Thank you to Valdone, who performed all of the western and Southern blots described in this thesis. To Jasmin, who carried out all of the MiSeq and whole genome sequencing bioinformatic analysis described. And to Andrew who managed all of my mouse lines, and helped with many mouse-handling experiments. Lastly, to Taka and Shantha, who have been a source of scientific wisdom all the way. I really appreciate all of your help. I would also like to thank Bryony who has been a constant source of great advice and support, especially in the more challenging times, both professional and personal!

I would also like to thank the ex-Division of Stem Cell and Developmental Genetics, and the current joint lab meeting group; including the Niakan lab and Lovell-Badge lab, for their great discussions and scientific input. Thank you to my thesis committee, Professor Abigail Tucker at King's College London, Professor Victor Tybulewicz, and Professor Robin Lovell-Badge, for helpful committee meetings, and PhD advice. And to Dr Peter Ellis, at the University of Kent, for his collaborative advice and input.

I would like to thank the Science Technology Platforms of the Crick who have provided extensive resources and expertise to allow me to discover many new avenues during my thesis project; the ASF, GeMS, BRF, BABS, Flow Cytometry and the EquipPark have contributed highly to this thesis. I would particularly like to thank GeMS who have

performed all of the ESC transfections to generate knock-ins, embryo microinjection and surgical procedures. The BRF, who have taken care of all animal husbandry. Also to Flow Cytometry who performed all of the ESC sorts.

I'm also grateful to the Communications and Engagement team that encouraged and supported my involvement in many 'extra-curricular' events and sparked my enjoyment for Science Communication. During my PhD I undertook an internship position, and I greatly appreciate the time that Fiona and all the team at the Science Media Centre took to teach me so much about their work and helped me think about my science from other perspectives.

Thank you to my friends; who always asked about my science and supported me throughout all of my university years. Thank you also to my PhD pals that have been there through thick and thin every step of the way.

To Matt and Scott, my flatmates and my best friends over many undergraduate and postgraduate years. I couldn't have done it without you. Living with you both over the past eight years has been the best and you're both family to me. Thank you to Jacob, who patiently learnt all about CRISPR-Cas9 genome editing, was excited as me when my 'cells were green' and who always believed in me. You have been a pillar of support throughout the final and hardest year, and I cannot express how grateful I am.

Thank you especially to Sam and to Mils, for their sound words of wisdom and constant encouragement. To Kate, who always provided great support and advice (and a great G&T). Lastly, and most importantly, the biggest of thank you's to my Mum and Dad; Heather and Nigel. I'm so incredibly grateful for everything you have done in helping me throughout my entire life and during my PhD. Your love and inspiration has always encouraged me to strive for more and become the person I am today. This thesis is dedicated to you. I hope you enjoy the reading.

Table of Contents

Abstract **3**
Acknowledgements **4**
Table of Contents **6**
Table of figures **9**
List of tables **11**
Abbreviations **12**
Chapter 1.Introduction **15**
1.1 Sex determination **16**
1.2 Evolution of the XY sex chromosomes **19**
1.3 X-chromosome functions in the germline **22**
1.3.1 Dosage compensation 25
1.3.2 X-chromosome upregulation 25
1.3.3 X-chromosome inactivation 27
1.3.4 Imprinted X inactivation in the mouse pre-implantation embryo 30
1.3.5 X-linked genes that escape XCI 33
1.4 The Y chromosome **33**
1.4.1 *Sry* 36
1.4.2 *Eif2s3y* 37
1.4.3 *Uty* 38
1.5 Development of CRISPR-Cas9 as a genome editing tool **39**
1.5.1 A brief history of CRISPR-Cas 39
1.5.1.1 Non-Homologous End Joining 42
1.5.1.2 Homology Directed Repair 44
1.5.2 Utilising CRISPR-Cas *in vitro* and *in vivo* to generate transgenic lines 46
1.5.3 Predicting CRISPR-Cas9 mutational outcomes 51
1.6 The requirement for specific sexes in the modern world – health, agriculture and pest control **52**
1.6.1 Layer hens 53
1.6.2 Dairy cows 55
1.6.3 Pest control 57
1.6.3.1 Gene drive 58
1.6.3.2 The disadvantages of gene drive 61
1.6.4 Skewing sex ratios using a bi-component system 62
1.7 Aims **64**
Chapter 2.Materials & Methods **65**
2.1 Mouse Lines **65**
2.1.1 Breeding and maintenance 65
2.1.2 Timed matings 65
2.1.3 Embryo collections 65
2.2 Molecular Biology **66**
2.2.1 Genomic DNA extraction 66
2.2.1.1 Mouse ear biopsies 66

2.2.1.2 Embryonic stem cells from 96-well plates 66
2.2.1.3 Embryonic stem cells or mouse tissue..... 67
2.2.2 Polymerase Chain Reaction (PCR) 67
2.2.3 MiSeq PCR 68
2.2.4 RNA extraction 68
2.2.5 cDNA synthesis.....69
2.2.6 Quantitative PCR (TaqMan probes).....69
2.2.7 TOPO XL cloning 69
2.2.8 Transformations and plasmid isolation.....69
2.3 Protein Biology 70
2.3.1 Protein extraction 70
2.3.2 Protein quantification 71
2.3.3 Western blot 71
2.4 Southern Blot.....71
2.5 Preparation of CRISPR Components.....73
2.5.1 sgRNA design, cloning and synthesis 73
2.5.2 Targeting vectors.....74
2.5.2.1 Cas9-eGFP 74
2.5.2.2 sgRNA-mCherry ("pLethal")..... 74
2.5.2.3 attB-sgRNA-mCherry-attB (TARGATT)..... 74
2.6 Embryonic Stem Cell Culture.....75
2.6.1 Preparation of cell culture plates.....75
2.6.2 Deriving embryonic stem cell lines.....75
2.6.3 Passaging and maintenance.....75
2.6.4 Freezing 76
2.6.5 Thawing 76
2.6.6 Fluorescence Activated Cell Sorting (FACS).....76
2.6.7 Standard transfections.....76
2.6.8 Generating knock-in lines by CRISPR-Cas9 homology directed repair (HDR) 77
2.7 Microinjections.....77
2.7.1 Zygote (TARGATT).....77
2.7.2 Blastocyst.....78
2.8 *In Vivo* Imaging 78
2.9 Next Generation Sequencing (NGS) 78
2.9.1 MiSeq library preparation.....78
2.9.2 MiSeq data analysis.....78
2.9.3 Whole genome sequencing (WGS) library preparation.....79
2.9.4 Whole genome sequencing data analysis.....79
2.9.5 RNAseq library preparation.....79
2.9.6 RNAseq data analysis.....80
Chapter 3.Results 1: Female specific lethality by expression of Cre inducible Diphtheria Toxin A 82
3.1 Introduction.....82
3.2 Results.....84
3.2.1 Timing of Cre recombinase expression 84

3.2.2 Cre recombinase-induced female-specific embryonic lethality 86
3.2.3 Effect of female lethality on litter sizes 89
3.2.4 Transcriptomic analysis of blastocysts 89
3.3 Discussion 96
Chapter 4.Results 2: Screening highly mutagenic sgRNAs to generate an sgRNA expressing knock-in mouse line 99
4.1 Introduction 99
4.2 Results 102
4.2.1 Deriving Rosa26 Cas9 eGFP embryonic stem cells 102
4.2.2 Candidate embryonic lethal gene: *Topoisomerase 1* 104
4.2.3 Generating the lethal sgRNA plasmid to screen sgRNAs 107
4.2.4 Generating mutations at *Top1* and evaluating mutation efficiency 108
4.2.5 Dynamics of sgRNA 1 CRISPR-Cas9 induced mutations 115
4.2.6 Generating the sgRNA 1 knock-in mouse model 118
4.3 Discussion 121
Chapter 5.Results 3: *In vivo* bi-component CRISPR-Cas9 induced mutations at *Top1* 127
5.1 Introduction 127
5.2 Results 128
5.2.1 Assessing mutations at the *Top1* locus in pre-implantation embryos in the bi-component CRISPR-Cas9 system 128
5.2.2 Evaluating the spectrum of mutations 130
5.2.3 Cas9/sgRNA embryos are non-viable 134
5.2.4 Effect of Cas9/sgRNA lethality on litter size 135
5.2.5 Effect on sex ratios 136
5.2.6 Post-implantation (E11.5) assessment of Cas9/sgRNA embryos 138
5.3 Discussion 140
Chapter 6.Results 4: Generating sex chromosome-linked Cas9 knock-ins 146
6.1 Introduction 146
6.2 Results 149
6.2.1 CRISPR-Cas9 HDR components for X-Cas9 HDR 149
6.2.2 Generating and characterising X-Cas9 embryonic stem cells 151
6.2.3 Assessing expression of X-Cas9 156
6.2.4 Assessing the bi-component system *in vitro* using X-Cas9 ESCs 158
6.2.5 Generating X-Cas9 chimeras and assessing germline transmission 161
6.2.6 CRISPR-Cas9 HDR components for Y-Cas9 HDR 164
6.2.7 Generating and characterising Y Cas9 embryonic stem cells 166
6.2.8 Producing Y Cas9 chimeras and assessing germline transmission 169
6.2.9 Investigating Cas9 expression by Cre recombinase removal of neomycin 173
6.3 Discussion 179
Chapter 7.Summary 183
Chapter 8.Appendices 191
8.1 Appendix A: Media and buffers 191
8.2 Appendix B: Oligonucleotides 197
8.3 Appendix C: TaqMan probes 200
8.4 Appendix D: Antibodies 200
Reference List 201

Table of figures

Figure 1. Evolution of the sex chromosomes..... 21
Figure 2. Mouse pre-implantation embryonic development..... 32
Figure 3. Schematic of CRISPR-Cas9 genome editing modes of producing mutations 42
Figure 4. Timeline of CRISPR-Cas technology development..... 48
Figure 5. CRISPR-Cas9 gene drive..... 60
Figure 6. Hemizygous X-Cre/Y male mating to a homozygous *R26*-DTA female..... 83
Figure 7. X-Cre induced TdTomato expression in female embryos..... 86
Figure 8. X-Cre induced Diphtheria toxin A female-specific lethality..... 88
Figure 9. Relative gene expression 91
Figure 10. *Hprt* and *Rosa26* relative gene expression..... 93
Figure 11. Differential gene expression analyses 95
Figure 12. CRISPR-Cas9 bi-component system for sex-specific lethality 100
Figure 13. Derivation of *R26*-Cas9 embryonic stem cells.....104
Figure 14. *Top1*/TOP1.....106
Figure 15. Generating the "pLethal" plasmid for screening sgRNAs 108
Figure 16. Low-throughput sgRNA screen.....115
Figure 17. Dynamics of sgRNA1 mutations.....117
Figure 18. Generating an sgRNA expressing mouse line by TARGATT knock-in 120
Figure 19. Evaluation of mutations at the *Top1* locus in pre-implantation embryos....130
Figure 20. Occurrence of a -1:1bp insertion mutation in individual embryos.....132
Figure 21. Developmental potential of transgenic sgRNA and Cas9 lines 137
Figure 22. Assessment of E11.5 embryos 139
Figure 23. Generating the components for X-Cas9 HDR 150
Figure 24. Characterising X-Cas9/Y ESCs.....153
Figure 25. Low-pass whole genome sequencing of X-Cas9 clones for karyotyping....155
Figure 26. X-Cas9 expression analysis 158
Figure 27. *Sry* sgRNA transfection and FACS/MiSeq in X-Cas9/Y ESCs.....160
Figure 28. Generating X-Cas9/Y chimeras and germline transmission.....163
Figure 29. Cas9 expression in a X-Cas9/X F1 generation female.....164
Figure 30. Generating the components for Y-Cas9 HDR 166
Figure 31. Generating a Y-Cas9 knock-in embryonic stem cell line.....168

Figure 32. Generating Y-Cas9 chimeras, germline transmission and assessing fertility 172
Figure 33. Transgene expression in X/Y-Cas9 and X-Cre/Y-Cas9 males 175
Figure 34. eGFP expression in X/Y-Cas9 males from X-Cre/Y-Cas9 parents 177
Figure 35. X/Y-Cas9 (-neo) matings to sgRNA/+ heterozygous females.....178

List of tables

Table 1. Occurrence of reads in Cas9/sgRNA embryos at E2.5 and E3.5 133
Table 2. Sequencing reads (low-pass WGS) for X-Cas9 ESC clones.....154
Table 3. Generating X-Cas9/Y chimeras.....162
Table 4. Sequencing reads (low-pass WGS) for the Y-Cas9 ESC clone. 169
Table 5. Generating the X/Y-Cas9 chimeras 170
Table 6. Hemizygous X-sgRNA/Y male mating to homozygous *R26*-Cas9 female 185
Table 7. Hemizygous X/Y-sgRNA male mating to homozygous *R26*-Cas9 female 185
Table 8. X-sgRNA(*Sry*) homozygous female mating with *R26*-Cas9 homozygous males 186
Table 9. Autosomal H11-sgRNA(*Sry*) homozygous female mating with *R26*-Cas9 homozygous males 187
Table 10. X-linked sgRNA male mating to hemizygous *R26*-Cas9 to generate wildtype offspring.....188
Table 11. Y-linked sgRNA mating with heterozygous *R26*-Cas9 to generate wildtype offspring.....188
Table 12. Embryo media.....191
Table 13. Cell culture media.....192
Table 14. DNA extraction buffers.....193
Table 15. Protein biology buffers.....195
Table 16. Southern blot buffers.....196
Table 17. Oligonucleotides.....199
Table 18. TaqMan probes.....200
Table 19. Antibodies 200

Abbreviations

6-tg	6-thioguanine
2i	Inhibitors (PD 0325901 and CHIR 99021)
ASF	Advanced sequence facility
BRF	Biological research facility
CAG	Cytomegalovirus (CMV) enhancer fused to chicken beta-actin promoter
Cbh	Cytomegalovirus (CMV) enhancer fused to chicken beta-actin promoter variant
cDNA	Complementary DNA
Cdx2	Caudal type homeobox 2
Cre	Cre recombinase
CRISPR	Clustered regularly interspersed short palindromic repeat
CTA	Cancer testis antigen
DAPI	4',6-diamidino-2-phenylindole
DSD	Disorders of sex development
DTA	Diphtheria toxin A
E	Embryonic day (age)
eGFP	Enhanced green fluorescent protein
EGFP	Green fluorescent protein
Eif2s3y	Subunit of eukaryotic initiation factor 2 (eIF2)
EGA	Embryonic genome activation
Epi	Epiblast
ESC	Embryonic stem cell
FACS	Fluorescence activated cell sorting
FISH	Fluorescence *in situ* hybridisation
FHM	Follicle holding medium
Gapdh	Glyceraldehyde 3-phosphate dehydrogenase
Gata6	GATA-binding factor 6
gDNA	Genomic DNA
GeMS	Genetic manipulation service

HDR	Homology directed repair
HMG	High mobility group
Hprt	Hypoxanthine guanine phosphoribosyl transferase
KSOM	Potassium simplex optimised medium
Indel	Insertion and/or deletion mutation
LIF	Leukaemia inhibitory factor
LINE	Long interspersed nuclear element
lssODN	Long single stranded oligonucleotide
MiSeq	Sequencing on the MiSeq platform
mCherry	Monomeric red fluorescent protein derivative
mRNA	Messenger RNA
MSY	Male-specific region on the Y chromosome
MYA	Million years ago
ncRNA	Non-coding RNA
Neo	Neomycin
NHEJ	Non-homologous end-joining
NLS	Nuclear localisation signal
OCT4	Octamer-binding transcription factor 4
PAM	Protospacer adjacent motif
PAR	Pseuoautosomal region
PCR	Polymerase chain reaction
PE	Primitive endoderm
PGK	Phosphoglycerate kinase
PN	Post natal day (age)
Pou5f1	Pou domain, class 5, transcription factor 1
qPCR	Quantitative real-time PCR
REC	Recognition lobe (Cas9)
RFP	Red fluorescent protein
RNAseq	RNA sequencing
Rosa26	ROSA beta geo 26
RT-PCR	Reverse transcriptase PCR
sgRNA	Single guide RNA

SHIMS	Single haplotype iterative mapping and sequencing
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SOX	SRY-related HMG box
SPRI	Solid phase reverse immobilisation
Sry	Sex determining region Y
TALEN	Transcription activator-like effector nucleases
TDF	Testis determining factor
TE	Trophectoderm
Tfrc	Transferrin receptor
Top1	Topoisomerase 1
Tsix	Gene antisense to *Xist*
TUB	Tubulin
Uty	Ubiquitously transcribed Y chromosomeY chromosome tetratricopeptide repeat protein
WGS	Whole genome sequencing
X:A	X chromosome to autosome ratio
Xa	Active X
XaXa	Two active X chromosomes
XAR	X added region
XCI	X chromosome inactivation/X-inactivation
Xi	Inactive X
XIC	X chromosome inactivation centre
Xist	X-inactive specific transcript
Xm	Maternal X chromosome
Xp	Paternal X chromosome
XUR	X chromosome upregulation/X-upregulation
YAC	Yeast artificial chromosome
YAR	Y added region

Chapter 1. Introduction

Sexual reproduction, the process of combining genetic information from two individuals into a single organism with a new genotype, is found in many species in the animal and plant kingdoms. The genetic information that will be inherited from the parents to the next generation is carried in a specific type of cell, called a gamete. In males, these gametes are called sperm. In females, the gametes are called oocytes. Fertilisation of an oocyte by a sperm results in the production of a zygote, the earliest form of the developing embryo, made up of just one cell. The zygote undergoes hundreds of cell divisions and cell differentiation processes to form a fully developed offspring.

The chromosome complement that the embryo inherited from the parents at fertilisation will also predispose the embryo to developing male- or female-specific characteristics. These male- or female-specific characteristics are known as sexual dimorphisms. Examples of sexual dimorphisms include anatomical or physiological differences, such as muscle mass, height and gonadal differentiation. Furthermore, sex differences are not limited to anatomy and physiology, but have also been shown to influence behaviour, personality and even susceptibility to some diseases (Snell and Turner, 2018). The sex-specific differences are largely dependent on which combination of sex chromosomes are inherited. The combination of sex chromosomes means that males and females inherit different chromosome complements and gene expression of these sex chromosomes induces downstream pathways to induce sex-specific primary and secondary characteristics.

The sex-specific characteristics seen in male/female anatomy and physiology have been exploited since the domestication of agricultural animals in order to maximise output for food and economic purpose. Examples of artificial selection include breeding male cows with larger muscle mass for meat consumption, or females with greater milk production. The growing human population size means there is an ever-increasing requirement for the surplus of a specific sex to increase production in farming and agriculture. Methods of generating single-sex litters to maximise output, for example in dairy cows or layer hens in agriculture, are actively being researched. Importantly, an understanding of the

regulation of genetic sex determination is required in order to influence Mendelian offspring sex ratios. The development of genetic tools to skew sex ratios is a major focus of this thesis.

In this thesis, I will first introduce the concept of genetic sex determination, the nomenclature of the mammalian sex chromosomes, and theory of sex chromosome evolution. I will then give further background to the heterogametic XY system in mammals; specifically, the role of the X and Y chromosomes in mammalian development, and sex chromosome gene expression. Finally, I will give a brief history to the development of CRISPR-Cas9 genome editing tools, and current methods of generating single-sex litters.

1.1 Sex determination

Around 335 BC, Aristotle stipulated that the temperature of the father during conception would determine the sex of the offspring produced; a father significantly warmer than the mother would result in a male child, whilst a cooler father would result in female child. Environmental theories regarding sex determination remained extremely well regarded for hundreds of years, and in fact, temperature remains an important regulator of sex determination in some reptile species to this day (e.g. Kohno et al., 2014). However, in the early 1900s, the ground-breaking discovery of the sex chromosomes by Nettie Stevens and Edmund Beecher Wilson changed the way we understand the genetics of sex determination. Stevens showed that male and female mealworms have two slightly different chromosome complements. Female mealworms carried 20 chromosomes, but males carried 19 plus 1 distinctly smaller chromosome. She described how this very small chromosome could be uniquely traced to sperm that produce only male offspring (Stevens, 1905b, Stevens, 1905a). This was the first description of the male-determining Y chromosome and showed that mealworm males are the ‘heterogametic’ sex. Heterogamety refers to the production of gametes which carry one of two possible sex chromosomes (as well as the autosomes). If the male is the heterogametic sex, such as in humans and mice, the sex of the offspring is dependent on which sex chromosome is

carried by the sperm that fertilises the oocyte. The homogametic sex produces gametes that all carry the same sex chromosome.

In 1909, Wilson concluded that inheritance of a specific combination of sex chromosomes was the genetic cause of sex determination (Wilson, 1909). He called the large chromosome "X" and the smaller "male" chromosome "Y". Therefore, males carried the sex chromosome complement XY. The X is inherited from the homogametic female that produces only X-carrying oocytes, and the Y from a Y-carrying sperm. Females carry the sex chromosome complement XX; an X from the oocyte and an X from a X-carrying sperm. This X/Y nomenclature is also used in the mammalian sex determination system. The nomenclature of the sex chromosomes was based on much earlier work by Hermann Henking in 1891 who called a large chromosome found in *Pyrrhocoris apterus* sperm "X". In 1960, an international panel determined that the autosomes would be numbered, and the X and Y sex chromosomes would remain lettered (Lejeune et al., 1960).

The X/Y nomenclature of the sex chromosomes is not used for all species. In birds, the heterogametic sex is the converse to mammals. Female birds have a sex chromosome complement ZW, producing two different sex chromosome-carrying gametes (Z or W). The male birds are the homogametic sex, with all gametes carrying a Z-chromosome. Male offspring therefore have the sex chromosome complement ZZ.

The early work by Stevens and Wilson confirmed that inheritance of a Y chromosome meant that the embryo developed as male. However it was unknown whether male development was directly caused by the Y chromosome, or what the influence of carrying a single X versus the two X chromosomes in females was. In 1959, the human male-determining factor was uncovered during two studies of human disorders of sex development (DSD); Turner syndrome (Ford et al., 1959) and Klinefelter syndrome (Jacobs and Strong, 1959). In both Turner syndrome and Klinefelter syndrome there is aneuploidy of the sex chromosomes. Turner syndrome patients have a single X chromosome (XO) whilst Klinefelter patients are trisomic, having two X chromosomes and a Y (XXY). Turner syndrome patients show female characteristics, despite having a single X whilst Klinefelter syndrome patients develop male characteristics, despite

having two X chromosomes. This finding highlighted that, in humans, the presence of the Y chromosome, and not the copy number of the X chromosome, gives rise to male characteristics. The testis-determining factor (TDF) therefore must be a Y chromosome-linked factor.

The first suggested TDF was a eutherian Y chromosome-linked gene called *ZFY* (Zinc finger protein Y-linked; Page et al., 1987), however this gene was autosomal in marsupials, suggesting it was unlikely to have a role in sex determination (Graves, 2006). Sinclair and colleagues also showed that *ZFY* was not expressed in the somatic cells of the testis, confirming that *ZFY* was unlikely to have a role in sex determination (Palmer et al., 1989). In 1990, a gene on the Y chromosome that encodes for male characteristics was discovered in humans. During a study of sex-reversed men, Sinclair *et al* discovered that the patients all had a translocation of the Y chromosome. Common to all of the patients was a small Y-linked gene contained within this translocation, called *SRY* (Sex-determining region on the Y; Sinclair et al., 1990). In the same year, the homologous *Sry* gene was discovered in the mouse (Gubbay et al., 1990). In 1991, experimental studies confirmed that *Sry* was an essential male sex-determining gene. Koopman *et al* showed that when a *Sry* transgene was introduced into female mouse embryos containing the usual complement of two X chromosomes, the females showed male developmental characteristics, such as testes (Koopman et al., 1991). Moreover, *SRY* was also found to be on the Y chromosome in marsupials (Foster et al., 1992). *SRY/Sry* is only present in eutherian and metatherian mammals. The platypus, a monotreme, has no *SRY* gene. They have five X chromosomes and five Y chromosomes, which are more similar to the bird ZW than to human XY sex chromosomes (Veyrunes et al., 2008). However, one potential TDF in protherians (e.g. platypus) is a Y-linked gene *AMH (*Anti-Mullerian hormone; Cortez et al., 2014), which is thought to have a role in sex-determination in many fish species (Hattori et al., 2012). The function of *SRY*/*Sry* in mammalian sex determination will be described further in 1.4.

1.2 Evolution of the XY sex chromosomes

Although the concept of genetic sex determination first arose in the early 1900s, it was not until some years later, in 1914, when Hermann Muller hypothesised how the sex chromosomes may have arisen during evolution, based on studies in *Drosophila*. Muller theorised that the sex chromosomes were unable to pair and undergo crossing-over during meiosis. Therefore the X and Y chromosome accumulated recessive deleterious mutations that were kept heterozygous due to a lack of recombination, and were therefore effectively neutral (Muller, 1914). Gradually, the number of recessive lethal mutations accumulated, resulting in the wildtype locus being lost from the population and the mutations become fixed. Muller hypothesised that the process of recombination and crossing-over was advantageous for the population and therefore sexual reproduction was preferable over asexual reproduction (Muller, 1914, Muller, 1964, Muller, 1932). The process of irreversible accumulation of deleterious mutations was coined "Muller's Ratchet" (Felsenstein, 1974).

In 1967, Susumo Ohno's hypotheses built on the concept that sex chromosomes originated from a pair of autosomes that originally could undergo normal pairing and recombination. Ohno suggested that a random mutation had occurred on one of the two autosomes, meaning the pair of chromosomes were then heteromorphic (Ohno, 1967). The gene containing this mutation is described as the early sex-determining gene and the chromosome became the "proto-Y" (grey line, Figure 1). The occurrence of this male-specific sex-determining gene on the proto-Y resulted in the surrounding proto-Y genes being selected for male-specific function (dark blue lines, Figure 1). The accumulation of male-function genes in this region inhibited recombination between the X and Y at this position (light blue, Figure 1). Ohno theorised that a gene duplication and inversion event of the Y-linked region containing the sex-determining gene further supressed crossing-over between the proto-X and proto-Y ("Ohno's hypothesis"; Ohno, 1967). The lack of recombination of the proto-X and proto-Y meant that deleterious mutations occurring at gene loci (red lines, Figure 1) were then lost by degeneration of the Y chromosome due to genetic drift. Over millions of years, this degeneration and loss of Y chromosome-

linked genes has led to a massive reduction in the Y chromosome size compared to the X, and the Y is extremely gene-poor (Figure 1).

The 'modern-day' XY sex chromosomes, evolving over the last 166 million years (Snell and Turner, 2018, Bellott et al., 2014); Figure 1), now encompass only a small region of continuous sequence homology between the two chromosomes, known as the pseudoautosomal region (PAR; purple, Figure 1). The PAR is thought to have arisen during an autosomal translocation event, as many of the genes contained within the PAR in eutherians are still autosomal in metatherians (Toder and Graves, 1998, Park et al., 2005, Waters et al., 2007). The human and mouse C57BL/6 PAR have very limited homology (Disteche et al., 1992) and humans have two separate PARs that lie at each end of the sex chromosomes (Cooke and Smith, 1986). In the mouse, the single PAR contains only five protein-coding genes (Mueller et al., 2013, Bellott et al., 2014). The presence of the PAR allows the X and Y chromosomes to pair during meiosis (Burgoyne, 1982, Ellis and Goodfellow, 1989). Deletion of the PAR results in failed X-Y pairing during meiosis and male infertility (Mohandas et al., 1992, Matsuda et al., 1992). The genes contained within the PAR are not susceptible to mechanisms of dosage compensation, as they are retained in two copies in males and females (Ellis and Goodfellow, 1989, Ross et al., 2005, Burgoyne, 1982).

To further understand the evolution of the modern X and Y chromosome, comparisons of gene content between three major groups of mammals; eutherians, metatherians and monotremes was performed. In all three groups, a region of the X chromosome is maintained, called the 'ancient' region. This conserved region contains homology to the chicken autosome 4 (Nanda et al., 1999, Ross et al., 2005, Bellott et al., 2010, Fridolfsson et al., 1998, Bellott et al., 2017). In eutherian mammals, the X chromosome has also gained additional autosomal regions fused to the X chromosome ("X-added regions", XARs), generating a larger X compared with marsupials and monotremes. This XAR contains homology to the chicken autosome 1 (Nanda et al., 1999, Ross et al., 2005, Bellott et al., 2010, Fridolfsson et al., 1998, Bellott et al., 2017). These extra regions have therefore arisen on the X chromosome since the divergence of eutherians and marsupials, approximately 166 MYA (Bellott et al., 2014). The XAR contains many genes shared by

the X and Y, suggesting that there was also additions to the Y chromosome ("Y-added region", YAR). Most likely, these additions also arose by translocations of autosomal regions and they have remained autosomal in marsupials and monotremes (Graves, 1995, Graves et al., 2006, Toder and Graves, 1998). Furthermore, modern-day mammalian and bird sex chromosomes have likely arisen as separate evolutionary events, shown by genomic and transcriptomic studies (Bellott et al., 2014, Bellott et al., 2010, Graves et al., 2006, Cortez et al., 2014). The chicken ZW sex chromosomes contain homology to the human autosomes 5 and 9, but not to the human sex chromosomes (Nanda et al., 1999).

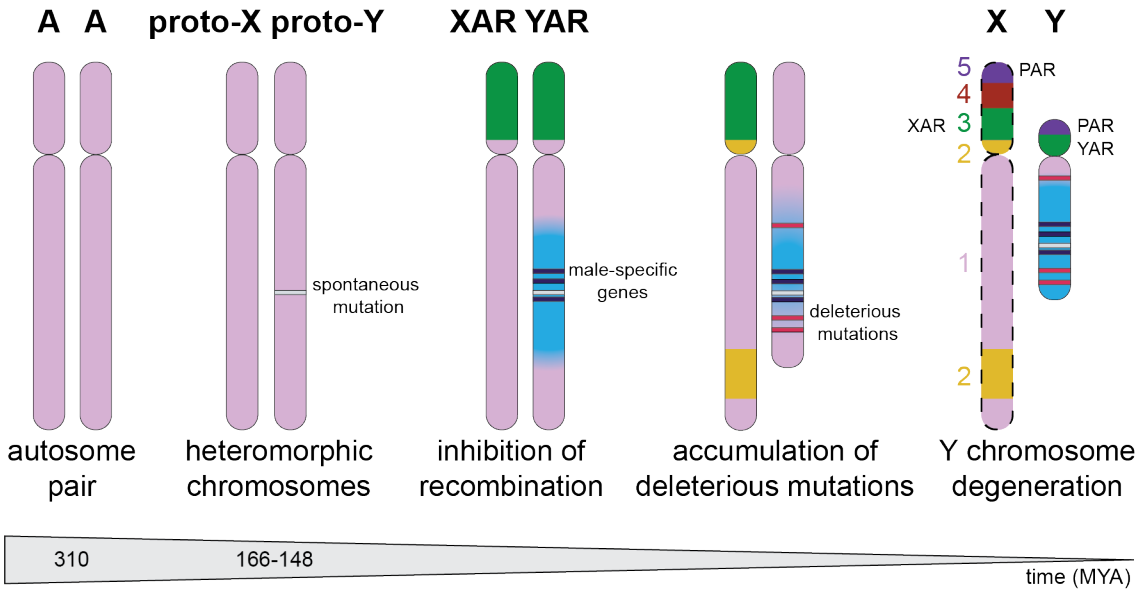

Figure 1. Evolution of the sex chromosomes

The modern X and Y sex chromosomes are thought to have arisen from an ancestral pair of autosomes approximately 166-148 million years ago (Snell and Turner, 2018, Bellott et al., 2014). One of these chromosomes, the proto-Y, acquired a spontaneous mutation (white line), in the early 'sex-determining' region. Accumulation of mutations on the sex specific proto-Y (dark blue lines) thereby resulted in inhibition of recombination between the proto-X and proto-Y. The proto-X is thought to have gained a new, fused region, translocated from an autosome (green); the X-added region (XAR). The inhibition of recombination resulted in the proto-Y gaining deleterious mutations (red lines) that are unable to be lost by recombination, thereby resulting in rapid degeneration of the proto-Y chromosome. Eventually, the Y chromosome was significantly degenerated, containing many multicopy, ampliconic, or male-specific genes. The modern day sex chromosomes contain only homology at the pseudoautosomal region (PAR; light purple). The X chromosome is thought to be made up of 5 evolutionary strata (coloured numbers).

Y-chromosome degeneration has resulted in retention of significantly fewer protein-coding genes compared to the X chromosome (Skaletsky et al., 2003). Given that so few

Y-linked genes remain, there has been much speculation about whether these persisting Y-chromosome genes have been retained by chance, or whether they have an essential male-specific function (Lahn and Page, 1997). This will be discussed further in 1.4. Another major consequence of the Y-chromosome degeneration was the unequal X-chromosome gene expression output between males that contain one X chromosome, and females that contain two X chromosomes. The system of female X chromosome inactivation (XCI) and X chromosome upregulation (XUR) ensures equal expression from the X chromosome in males and females, and in balance with diploid autosomes. The X chromosome and autosome balance is called "dosage compensation" and was hypothesised by Susumo Ohno (Ohno, 1967). Dosage compensation is described in greater detail in 1.3.

1.3 X-chromosome functions in the germline

Autosomal genes spend an equal amount of time in males and females, as both of the sexes are diploid. Therefore, if a mutation arises, any recessive effects can be neutral in a heterozygous genotype. However for the sex chromosomes it is more unbalanced. When the male is the heterogametic sex, such as in humans, the X chromosome is present two-thirds of the time in females, but only one-third of the time in males (Rice, 1984). If a mutation arises on the hemizygous X chromosome in males, there is no wildtype allele to mask the mutated allele, and this mutation could result in a fitness benefit or cost. If the effect is beneficial to the male, it may be fixed and spread through the population. If the effect has a fitness cost, it may be deleterious and lost in the population. Therefore, recessive mutations on the X chromosome may be under greater selection compared to the autosomes. This comparatively increased rate of selection on the X is called the "faster-X" effect (Vicoso and Charlesworth, 2009, Vicoso and Charlesworth, 2006).

The "faster-X" may have a higher rate of selection acting on the chromosome compared to autosomes when recessive mutations occur, but this rate can also be influenced by mutations in sex-biased genes. If a sex-biased autosomal gene gains a mutation that provides a fitness benefit for females, but not for males, it will only increase in the population if the advantage is significantly higher than the disadvantage. However if the

mutation occurs in an X-linked gene that provides a fitness benefit for females, it is more likely to be passed on to the next generation, because the X chromosome is present more often in females (two-thirds versus one-third in males). On the converse, X-linked mutations that are beneficial for males but have a fitness cost for females, can also become fixed quickly, as the fitness effect is expressed in single X-carrying males, and is hidden by heterozygosity in females. Overall, there is a sexual antagonism between sex-bias genes. The X chromosome can accumulate these male-specific or female-specific beneficial mutations, resulting in an accumulation of sexual antagonism of sex-specific genes carried on the X chromosome relative to the autosomes. This is known as Rice's hypothesis (Rice, 1984, Vicoso and Charlesworth, 2006).

Although the X is present more in females, many male-specific genes on the X chromosome are exclusively expressed in pre-meiotic and/or post-meiotic germ cells in the testis (Hurst, 2001, Wang et al., 2001, Lercher et al., 2003, Khil et al., 2004). A screen for spermatogonial-expressed genes in the mouse showed that a disproportionately high number are X chromosome-linked (Wang et al., 2001). Of the genes expressed in pre-meiotic germ cells, many are single copy (Wang et al., 2001). Conversely, many of the testis-specific genes expressed in post-meiotic germ cells are highly ampliconic, referring to duplications of regions that are greater than 10kb in length and share greater than 99% sequence similarity. These ampliconic genes contribute approximately 13% of the X chromosome in humans, and 17% in the mouse (Mueller et al., 2013, Mueller et al., 2008, Ross et al., 2005). Many of the X-linked amplicon genes are cancer testis antigens (CTAs). CTA genes are highly expressed in multiple cancer types and in spermatogonia (Stevenson et al., 2007, Simpson et al., 2005). Many of the CTA gene families on the X chromosome are organised into complex inverted repeats (Simpson et al., 2005). As many as 10% of genes on the human X chromosome are in CTA families (Ross et al., 2005).

During male germ cell meiosis, the X chromosome (and Y chromosome) are silenced, in a process called meiotic sex chromosome inactivation, MSCI (Turner, 2007). It has therefore been suggested that the post-meiotic germ cell genes have been amplified on the X chromosome in order to achieve sufficient expression levels post-MSCI, when the majority of X-linked genes remain silenced (Mueller et al., 2008, Turner, 2007), and that

the amplified X-genes may be essential for spermiogenesis and sperm elongation. Further, in the mouse, functional pseudogenes have been copied onto autosomes from the X chromosome by retrotransposition (Adra et al., 1988). Transposition onto an autosome may allow escape from MSCI. Unlike single copy and multi copy genes, ampliconic genes do not share homology between humans and mice and appear to have been acquired independently (Mueller et al., 2013).

Furthermore, X-linked genes that have testis-specific function have also been shown to be expressed very highly in the brain compared to other somatic tissues (Zechner et al., 2001, Nguyen and Disteche, 2006). Over 100 X-linked genes have been implicated in patients with intellectual disabilities (Skuse, 2005, Ropers, 2010).

The X chromosome also contains genes that have essential female functions. As earlier described, the X chromosome consists of 'ancient' regions that are shared by eutherians and metatherians, and 'added' regions which are unique to eutherians. Genes contained within the ancient region of the human and mouse X chromosome are usually female-biased and highly expressed in the ovary and placenta (Khil et al., 2004). The ovary-specific X-linked genes are dosage sensitive; XO (Turner syndrome) females who have a single X present ovarian dysgenesis, supporting theories that these ovary-specific X-linked may be required in two copies.

Rice's hypothesis highlights that sexual antagonism may allow genes that have a role in male- or female-specific function to become accumulated, either on the X chromosome or via translocation to an autosome. This is a different evolutionary force to Ohno's hypothesis, which predicted equal dosage of sex chromosome-linked genes between males and females, thereby suggesting that gene amplification would be unlikely. Therefore, X-linked genes may be under different evolutionary forces dependent on gene function.

1.3.1 Dosage compensation

Females carry two X chromosomes, whilst males have a single X. Therefore, the X-linked protein-coding genes are present in twice the copy number in females, compared to males. Ohno hypothesised that a mechanism of dosage compensation is employed to balance the X chromosome output between males and females and prevent 'aneuploidy-like' effects in male cells (Ohno, 1967). The unbalanced gene dosage in males could be simply combatted by up-regulating all X-linked genes two-fold, to match the X chromosome output of female cells. However, a consequence of the two-fold X-linked gene upregulation is a two-fold increase in X chromosome to autosome output in females. This problem of dosage is thereby solved in 'two-steps'. Firstly, X-linked genes in male and female cells are up-regulated (X-upregulation; XUR), resulting in an X chromosome to autosome (X:A) ratio of '1' in males, and '2' in females. The output of the X chromosome is therefore balanced with the diploid autosomes, in males. This process of XUR will be described in further detail in 1.3.2. In females, the second step, X chromosome inactivation (XCI) occurs. In this process, one X chromosome is transcriptionally silenced, bringing the overall X chromosome gene expression level back to '1', equivalent to the autosomes. This will be discussed further in 1.3.3 and 1.3.4.

1.3.2 X-chromosome upregulation

Ohno hypothesised that the single male X chromosome would need to balance with autosomal gene output, by upregulating X-linked genes two-fold. The extent of XUR is calculated by dividing the medium expression of X-linked genes (X) by the medium expression of autosomal genes (A); giving the X:A ratio. An X:A=1 indicates that the X-linked genes have been upregulated, in keeping with Ohno's hypothesis. An X:A=0.5 suggests that the X-linked gene output has not been upregulated and is half the autosome output. The X:A ratio also takes into account the number of active X chromosomes compared to the autosomes. Lyon and Grumbach hypothesised that cells are able to 'count' the number of active X chromosomes (Lyon, 1962, Grumbach et al., 1963). Cells are therefore able to maintain a single active X chromosome with the expected X:A ratio,

even if the females have a single X (XO, Turner syndrome) or have multiple Xs (e.g. XXY, Klinefelter syndrome).

Evidence for XUR first arose from analysis of the gene *Clcn4*. In *Mus musculus* this gene is autosomal, whilst in a different mouse species, *M. spretus*, the gene is X-linked. Adler *et al* showed that expression of *Clcn4* in *M. spretus* is twice that of *M. musculus*. The two-fold up-regulation of *Clcn4* in *M. spretus* was the first indication that X-linked genes undergo XUR (Adler et al., 1997). Subsequently, XUR has been assayed at the whole chromosome level. Two studies utilised microarrays to support Ohno's hypothesis of XUR (Gupta et al., 2006, Nguyen and Disteche, 2006). However an RNAseq study reported the opposite; with X:A ratio of approximately 0.5, thereby rejecting Ohno's hypothesis (Xiong et al., 2010). A reanalysis of this RNAseq data was performed more recently, filtering out non- and lowly-expressed genes. These original RNAseq studies by Xiong *et al,* were performed in somatic tissue, and therefore highly enriched and ampliconic X-linked genes which are only expressed in the testis and ovary were included in the analysis. The inclusion of these lowly-expressed genes artificially lowered the X:A ratio. When the lowly expressed genes were excluded in the reanalysis, the X:A ratio confirmed XUR (X:A~1) and upheld Ohno's hypothesis (Deng et al., 2011, Kharchenko et al., 2011, Lin et al., 2011, Yildirim et al., 2011). Sangrithi *et al* also showed that in somatic tissue, when all genes were included, the X:A ratio were below one. When the authors removed genes that were lowly expressed, the X:A ratio equalled at least one, showing that male and female somatic tissue undergo XUR (Sangrithi et al., 2017). Studies showed that XUR preferentially affects genes with housekeeping function and X-linked gene products that interact with autosomes (Pessia et al., 2012). The expression and dosage of these genes therefore needs to be tightly regulated (Birchler, 2012, Pessia et al., 2012, Pessia et al., 2014, Sangrithi et al., 2017, Sangrithi and Turner, 2018).

Other studies have performed slightly different analyses, for example by comparing the ratio of the modern X chromosome to the proto-X using human and chicken chromosomes (Lin et al., 2012), or by only comparing protein-coding genes that are known to be dosage-sensitive (Pessia et al., 2014). These studies confirmed that XUR affects X-linked genes that encode protein complexes that interact with autosomal genes.

By comparing modern X-linked genes that have autosomal orthologues in the chicken, these studies also showed that there is downregulation of autosomal genes that interact with X-encoded proteins (Julien et al., 2012, Lin et al., 2012, Pessia et al., 2012). In 2019, studies in bovine embryos, which undergo random XCI at the blastocyst-stage (Bermejo-Alvarez et al., 2011) confirmed evidence for XUR with X:A ratios of one in somatic and germline cells (Duan et al., 2019, Ka et al., 2016, Jiang et al., 2014).

1.3.3 X-chromosome inactivation

XCI reduces genetic output of the two X chromosomes in females from two to one by transcriptionally silencing a single X. Major breakthroughs in understanding XCI came during the 1960s. Mary Lyon observed that female mice heterozygous for a mutation in an X-linked gene show variability in coat colour. She speculated that these patches of different colour had arisen as a clonal growth of a specific type of cell that had undergone silencing of a single X chromosome (Lyon, 1961). Lyon's Law is the first description of dosage compensation by X chromosome silencing in females. Russell and Bangham also showed that similar coat colour patterns occur in mice with autosomal translocations of the heterozygous X chromosome coat colour loci (Russell and Bangham, 1961).

Lyon theorised that XCI occurs in three stages; initiation, spreading and maintenance (Lyon, 1988) and that XCI may be initiated at a specific region of the X chromosome, termed the X-inactivation centre (XIC). Rastan and Robertson (1985) showed that two XIC X-linked regions must be present in order for random XCI to occur (Rastan and Robertson, 1985). This finding suggested that the XIC had a role in 'choosing' which X chromosome is silenced prior to the onset of XCI. As well as the process of 'counting' the number of X chromosomes by the X:A ratio, it is thought that the XIC may also function in 'sensing' (Augui et al., 2007a). Sensing is very similar to counting but takes into account the potential of extra X chromosomes to undergo XCI.

In the 1990s, many studies utilised translocations of the X chromosome to pinpoint the exact location of the X-linked XIC (Brockdorff et al., 1991, Leppig et al., 1993, Kay et al., 1993). Once the XIC region was narrowed down, a candidate gene for XCI was

discovered in 1991. Brown *et al* discovered a non-coding RNA (ncRNA) that was found to have female-specific expression in human somatic cells (Brown et al., 1991, Brown et al., 1992) and was situated within the XIC. The ncRNA was called *XIST* (X-inactive specific transcript). The following year, the mouse homologue *Xist* was discovered (Borsani et al., 1991, Brockdorff et al., 1992, Kay et al., 1993). Studies utilising RNA fluorescence *in situ* hybridisation (RNA-FISH) showed that the *XIST*/*Xist* RNA molecule specifically coats the inactive X (Xi) and can be detected as an RNA cloud in the nucleus (Clemson et al., 1996, Brown et al., 1992, Engreitz et al., 2013, Simon et al., 2013). To elucidate how *XIST*/*Xist* may function to silence the inactive X chromosome, Penny *et al* deleted a region of one of the *Xist* alleles in female embryonic stem cells (ESCs) and proved that *Xist* is an essential requirement of XCI. Further, targeting *Xist* in ESCs resulted in an XCI skew, inducing *Xist* expression from the non-targeted allele (Penny et al., 1996). An *Xist* cDNA transgene integrated onto an autosome induced silencing of the autosomal genes, showing that *Xist* was functional to silence genes that were not X-linked (Wutz and Jaenisch, 2000). Lee *et al* utilised a yeast artificial chromosome (YAC), where the *Xist* gene was contained within a 450kb region. The YAC was transfected into XY mouse ESCs. Upon differentiation of the ESCs, results showed that the *Xist* transgene was able to induce silencing in many of the genes in the surrounding YAC region (Lee et al., 1996, Lee and Jaenisch, 1997). These results confirmed *Xist* as an essential gene for XCI.

Activation and expression of *Xist* generates a ncRNA which is spliced and polyadenylated before coating the Xi *in cis* (Brockdorff et al., 1991, Borsani et al., 1991, Clemson et al., 1996, Brown et al., 1992, Engreitz et al., 2013, Simon et al., 2013). To investigate some of the other factors regulating *Xist* expression, studies have generated loss-of-function mutations in candidate genes and assessed the effect on XCI (Pontier and Gribnau, 2011). There are several X-linked genes that are proposed to have a role in *Xist* activation. One example is the X-pairing region (*Xpr*) which mediates interactions between the two XICs in XX cells (Augui et al., 2007b). Another example is *Rnf12*/RLIM (Ring finger protein, LIM domain interacting), a protein-coding gene (Jonkers et al., 2009) which is thought to regulate X chromosome counting and activate *Xist* directly. Overexpression of *Rnf12* in either male or female cells triggers ectopic XCI (Jonkers et al., 2009), and *Rnf12* knock-

out in females shows reduced XCI (Barakat et al., 2011). A third example is *Jpx (Enox)*. *Jpx* (Just proximal to Xist*)* encodes for a ncRNA and may enable *Xist* expression by interfering with *Tsix* (X-inactive specific transcript antisense; Tian et al., 2010, Jonkers et al., 2009). *Tsix* is an anti-sense RNA that represses *Xist* activity (Lee et al., 1999) thereby maintaining the transcriptional activity of the active X (Stavropoulos et al., 2001, Shibata and Lee, 2004, Sado et al., 2006, Ohhata et al., 2006). Deletion of the *Tsix* promoter in one allele of XX cells *in vitro* results in skewed XCI (Lee and Lu, 1999, Lee, 2002, Migeon et al., 2002, Stavropoulos et al., 2001) but does not result in ectopic *Xist* upregulation (Morey et al., 2001). A homozygous deletion of *Tsix* however, does appear to result in increased *Xist* expression (Lee, 2005) and the authors suggest *Tsix* has a role in X-chromosome counting. Once *Xist* is expressed, it is thought to recruit chromatin remodelling complexes to the Xi. One example of these proteins is PRC2 (Polycomb repressive complex 2), which trimethylates lysin 27 on histone H3 (H3K27; Marks et al., 2009, Marks et al., 2015), an epigenetic mark associated with inactive chromatin. The XIC region upstream of *Xist,* containing *Jpx* and *Xpr,* amongst others, is also highly enriched in H3K9 dimethylation and H3K27 trimethylation (Heard et al., 2001, Rougeulle et al., 2004).

Once *XIST*/*Xist* is expressed, the ncRNA coats the X chromosome in *cis*. Lyon proposed that "long interspersed nuclear elements" (LINEs), which are highly repetitive elements abundant on the mouse X chromosome, may facilitate *Xist* spreading (Lyon, 1998). Bailey *et al* showed that the mouse X chromosome is enriched in LINEs compared to the autosomes, and that there is significant clustering of these elements around the XIC (Bailey et al., 2000). They hypothesised that this enrichment of LINEs on the X chromosome may allow for propagation of XCI along the chromosome (Bailey et al., 2000). Chromosome silencing due to *Xist*-containing translocations are significantly less efficient at *Xist* spreading compared with on the X chromosome. This suggests that these LINE-poor regions have slower *Xist* spreading (Sharp et al., 2002). Ross *et al* also showed that the human X chromosome was significantly enriched for LINEs. They confirmed that the XIC was highly LINE rich, although interestingly the *Xist* locus was comparatively LINE poor (Ross et al., 2005).

1.3.4 Imprinted X inactivation in the mouse pre-implantation embryo

After fertilisation of the mouse oocyte to form the single-cell zygote embryo, successive rounds of cell division give rise to the 2-cell (E1.5), 4-cell (E2.0) and 8-cell (E2.5) stages of the pre-implantation embryo. At the 2-cell stage, embryonic genes become transcriptionally active, in a process called embryonic genome activation (EGA; Flach et al., 1982, Bernstein and Mukherjee, 1972). Prior to EGA, embryonic development relies on a store of mRNAs deposited in the embryo by the maternal oocyte. The maternally-loaded mRNAs are degraded after EGA (Clegg and Piko, 1983, De Leon et al., 1983, Wassarman and Kinloch, 1992). After the 8-cell stage, the embryo undergoes compaction and asymmetric divisions, giving rise to the morula-stage (16-32 cells, E3.0) embryo. Individual cells of the morula-stage embryo are morphologically indistinct (Ducibella and Anderson, 1975). The totipotent blastomeres (individual cells of the embryo) become lineage specified as the embryo develops, due to the expression of different factors. By E3.0 to E3.5, the late-morula to early-blastocyst-stage, the blastomeres have been lineage specified to form either the inner cell mass (ICM) or the trophectoderm (TE; Figure 2). This lineage specification is the first cell-fate decision. By E4.0-4.5, the late blastocyst-stage, the embryonic cells have further diverged into three lineages (Gardner, 1985, Dietrich and Hiiragi, 2007, Rossant and Tam, 2009). Cells of the ICM become either epiblast (Epi) or primitive endoderm (PE, Figure 2). The Epi is marked by expression of transcription factor *Nanog* (Chazaud et al., 2006, Chambers et al., 2003, Mitsui et al., 2003), whilst the PE is marked by expression of *Gata6* (GATA-binding factor 6; Chazaud et al., 2006). The pluripotency factor OCT4 (Octamer-binding transcription factor 4, encoded by *Pou5f1;* POU class 5 homebox 1) is highly expressed in ICM cells, maintained in Epi cells, and downregulated in PE cells (reviewed in e.g. Chazaud and Yamanaka, 2016). Cells of the Epi are destined to become the embryo-proper, whilst cells of the PE will become the yolk sac. The trophectoderm (TE) cells, marked by *Cdx2* (Caudal type homeobox 2) expression (Strumpf et al., 2005, Ralston and Rossant, 2008, Niwa et al., 2005) surround the blastocyst (Figure 2) and will form the extra-embryonic, placental tissues.

During the period of mouse pre-implantation development prior to the first cell-fate decision, XCI is imprinted in all blastomeres, i.e. an X chromosome is inactivated in a

parent-of-origin specific manner. The imprinting results in the paternally-inherited X chromosome (Xp) always being transcriptionally silenced in XX female mouse embryos (Takagi and Sasaki, 1975, Okamoto et al., 2004, Okamoto et al., 2005). The phenomenon of imprinted Xp XCI is not unique to pre-implantation mouse embryos, but also occurs in the somatic cells of metatherian mammals (Sharman, 1971). Mouse pre-implantation Xp XCI was originally thought to be a result of retained Xp silencing after MSCI (Huynh and Lee, 2003). However MSCI is *Xist*-independent (Turner et al., 2002, Turner et al., 2006) and there is Xp transcriptional activity at the 2-cell stage, at the onset of EGA (Okamoto et al., 2005). A 1997 study by Marahens *et al* showed that if *Xist* is deleted from the paternal X chromosome, females were severely underdeveloped. However if *Xist* is deleted from the maternal X, females developed normally (Marahrens et al., 1997) confirming the role of *Xist* in imprinted Xp XCI. Furthermore, *Xist* was shown to be upregulated at approximately the 2-4 cell stage, closely after XCI is initiated, supporting the role of *Xist* in Xp XCI (Okamoto et al., 2005, Okamoto et al., 2004, Zuccotti et al., 2002). Nuclear transfer studies had suggested that imprinted XCI is activated by a repressive mark that silences maternal *Xist,* established during oogenesis (Tada et al., 2000). Key studies were published in 2017, when Inoue *et al* discovered that the epigenetic mark H3K27me3 was responsible for autosomal imprinting (Inoue et al., 2017a). A follow-up study again by Inoue *et al* showed that the *Xist* locus is coated with H3K27me3 marks during oogenesis. The H3K27me3 repressive mark is maintained throughout pre-implantation development on the maternal X chromosome, and loss of H3K27me3 allows ectopic Xm *Xist* expression (Inoue et al., 2017b, Inoue et al., 2017a).

At the blastocyst-stage, E4.0-4.5, the paternally-silenced X chromosome in the Epi is reactivated and *Xist* coating is lost (Mak et al., 2004, Okamoto et al., 2004). This leads to a brief window during embryonic development when cells of the Epi contain two active X chromosomes (XaXa, (Mak et al., 2004), Figure 2). Following this period of XaXa in the blastocyst Epi, the cells of the Epi undergo random XCI mediated by *Xist,* similarly to XCI in humans and rabbits (Okamoto et al., 2011). Conversely, cells contributing to the TE and PE that form the extra-embryonic lineages maintain imprinted XCI, and the Xp remains transcriptionally silenced (Takagi and Sasaki, 1975), (West et al., 1977); Figure 2).

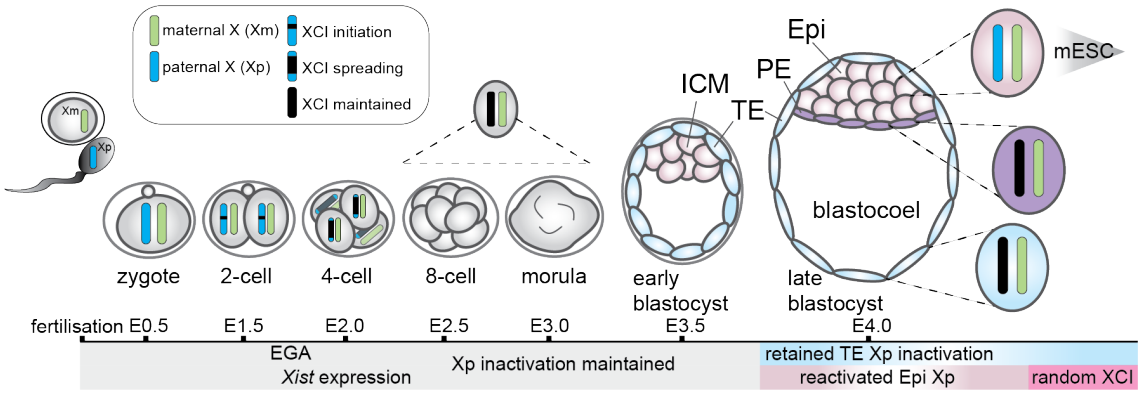

Figure 2. Mouse pre-implantation embryonic development

After fertilisation of the X-carrying oocyte (maternal X; Xm) by an X-carrying sperm (Xp), a female (XX) zygote is generated. XCI in the pre-implantation mouse embryo is imprinted. *Xist* is expressed from the Xp and spreads *in cis* along the Xp, transcriptionally silencing the chromosome. This XCI is maintained up until the early blastocyst-stage. At the later blastocyst-stage both X chromosomes are briefly reactivated in the epiblast (Epi). This cell type can be used for generating embryonic stem cells. In the primitive ectoderm (PE) and trophectoderm (TE), imprinted Xp silencing is maintained. The cells of the epiblast then undergo random XCI.

Although XCI in mice is imprinted, specifically only silencing the paternal X chromosome in pre-implantation embryos, and trophectoderm-derived lineages, does not occur in other mammals. In humans and rabbits there is no imprinted XCI but is instead random in early embryonic development (Okamoto et al., 2011). Whether random or imprinted however, XCI has been observed in all mammalian species studied to date (Ohno et al., 1959, Lyon, 1961, Heard et al., 1997).

In summary, XUR and XCI are essential during development for balanced X-chromosome gene dosage between males and females. However the mechanism of XUR and XCI is an important consideration when generating genetic tools. For example, transgenes on the X chromosome may be susceptible to XUR and XCI. This is of particular note when generating paternal X-linked transgenic mouse models, as they may be subject to Xp XCI. The generation of X-linked transgenes will be discussed in greater detail in Chapter 3 and Chapter 6.

1.3.5 X-linked genes that escape XCI

XCI balances gene dosage between males and females. However there are a number of X-linked genes that 'escape' XCI, and are not transcriptionally silenced.

In 1962, Mary Lyon suggested that genes present in the PAR would escape XCI (Lyon, 1962). While this has been confirmed (Berletch et al., 2010, Berletch et al., 2011, Disteche et al., 2002), it is now known that the X-linked escapees are not limited to the PAR (Berletch et al., 2010, Berletch et al., 2011). Up to 3% of genes on the mouse X chromosome consistently escape XCI (Berletch et al., 2011), with a further 4% presenting expression patterns representative of XCI escape in a tissue-specific manner (Carrel and Willard, 1999, Yang et al., 2010, Disteche and Berletch, 2015, Berletch et al., 2015, Tukiainen et al., 2017, Garieri et al., 2018). In the mouse, escape genes are located across the entire X chromosome, but the majority appear to be localised to the XAR. These escape genes lack *Xist* coating and appear to lack repressive histone marks that are characteristic of silenced X-linked genes e.g. H3K27me3 (Simon et al., 2013, Goto and Takagi, 2000, Yang et al., 2010, Marks et al., 2015). In the human, there are a higher number of escapees compared to the mouse, and the escapees are more highly concentrated to the human XAR. Approximately 15% of X-linked genes consistently escape XCI in humans, with a further 10% escaping XCI in a tissue-specific manner and are highly diverse (Carrel and Willard, 2005). Escapee genes are hypomethylated at CpG islands compared to non-escape genes and this CpG island hypomethylation has been used to identify further putative escapees (Lister et al., 2013, Cotton et al., 2015, Schultz et al., 2015).

1.4 The Y chromosome

The degeneration of the Y chromosome has led to significantly less protein-coding genes on the Y than the X chromosome (Bellott et al., 2014). In humans, only 3% of ancestral genes have survived (Bellott et al., 2010, Skaletsky et al., 2003), whereas 98% of X chromosome genes have remained (Mueller et al., 2013). The degeneration of the Y chromosome gained traction in mainstream media; citing the predicted extinction of

males due to the loss of all protein-coding genes on the Y chromosome in the next 5-10 million years (Graves, 2004, Aitken and Marshall Graves, 2002). The concept of the "impeding demise" hypothesis is that the human Y chromosome would lose approximately five protein-coding genes that originally had X-homologues every one million years (Graves, 2004, Aitken and Marshall Graves, 2002). In 2005, Hughes *et al* showed that this was not the case. They performed comparative sequencing analyses between chimpanzee and human, which diverged approximately six million years ago. The impending demise hypothesis would predict that the chimpanzee should have many more Y chromosome linked genes that have since been lost in the human. Their findings contradicted the impending demise hypothesis and showed that there had been very little Y-gene loss during the six million years of human evolution (Hughes et al., 2005). Further studies by Hughes *et al* again showed that although Y-decay was initially extremely rapid, the degeneration since the divergence of humans has practically stopped, and the Y chromosome has undergone recent acquisition and amplification of testis-specific genes (Hughes et al., 2012, Hughes et al., 2010). Studies by Bellott *et al* and Cortez *et al*, confirmed that the Y chromosome gene loss had practically stopped since the divergence of monkeys, chimps and humans (Bellott et al., 2014, Cortez et al., 2014). It is thought that the genes retained on the Y chromosome have essential biological functions, such as transcription, translation and protein stability (Bellott et al., 2014, Lahn and Page, 1997).

The degeneration and divergence of the Y chromosome resulted in a large region of the Y that is unable to pair with the X, known as the male-specific region on the Y chromosome, MSY (Skaletsky et al., 2003). In humans, the MSY is flanked by the two PARs, where pairing and meiotic recombination normally occurs. The MSY contains highly ampliconic genes and repetitive elements, preventing accurate sequencing of the region. In 2003, Skaletsky *et al* performed the first sequencing of the human MSY by "single-haplotype iterative mapping and sequencing" (SHIMS). Their data showed that the MSY is made up of three main types of euchromatic sequences. Firstly, X-transposed units; which are highly sequence-similar to X chromosome genes. Secondly; X-degenerate genes, which occurred on the ancestral autosomes and degenerated during sex chromosome evolution. Lastly, ampliconic genes, which contribute approximately 30% of the MSY euchromatic sequence, and are multi-copy gene families. Of the ampliconic

gene families, two are thought to have originated on the early Y chromosome, whilst seven are thought to have been transposed from autosomes. Of the nine ampliconic gene families, at least six are proposed to have a role in testis-function (Skaletsky et al., 2003). The SHIMS method of sequencing and mapping the complex Y chromosome has since been used to evaluate primate Y, human X and chicken Z chromosomes (Bellott et al., 2014, Bellott et al., 2010, Skaletsky et al., 2003, Mueller et al., 2013, Hughes et al., 2012, Hughes et al., 2010, Hughes et al., 2005) and ancestral regions of the Y chromosome in eight mammalian species for comparative studies (Bellott et al., 2014).

The most extensive sequencing and annotation of the mouse MSY was performed in 2014 by Soh *et al* (Soh et al., 2014). The mouse MSY has undergone extensive degeneration from loss of ancestral genes, and gain of genes by transposition from the autosomes. The study showed that the mouse MSY is highly euchromatic (89.5Mb, 99.9%). This is contrary to suggestions that the mouse Y chromosome would be very small and heterochromatic (Bachtrog, 2013, Charlesworth and Charlesworth, 2000, Graves, 2006). In the 89.5Mb of euchromatic sequence, Soh *et al* described two types of genes; ancestral Y-genes and acquired Y-genes. The ancestral genes made up 2Mb of the euchromatin, and originated from the ancestral autosomes. The remaining 87.5Mb consisted of acquired sequence and was almost entirely ampliconic. Of this 87.5Mb of acquired sequence, the majority (86.4Mb) was located on the Y chromosome long arm (Yq). Yq amplicons fall into three large gene families; *Sly* (Sypc3-like Y-linked), *Srsy* (Serine-rich, secreted, Y-linked) and *Ssty* (Spermiogenesis specific transcript on the Y; Soh et al., 2014). On the Y chromosome short arm (Yp), there are twelve families of protein-coding genes, including *Sry* (Sex-determining region Y), *Uty* (Ubiquitously transcribed tetratriopeptide repeat containing, Y linked) and *Eif2s3y* (Soh et al., 2014). These three genes in particular will be discussed in greater depth. *Sry* and *Eif2s3y* have essential functions in fertility, and may be useful targets for generating single sex litters in the future. *Eif2s3y* and *Uty* are expressed in embryonic stem cells and therefore provide useful Y-chromosome expression markers throughout this thesis.

The results of the Soh *et al* study showed that the mouse MSY has greatly diverged from the human and primate MSY through the huge amplification of three gene families. Only

2Mb (2.2%) of ancestral shared genes are retained on the MSY (Soh et al., 2014). The highly-amplified gene families have a specific role in male germ cell development. A reduction in copy number of these multicopy genes results in a sex skew, producing more females than males (Conway et al., 1994). The same effect is also seen upon knock-down of *Sly* (Cocquet et al., 2009, Cocquet et al., 2012). The authors suggest therefore that the amplification of these gene families on the MSY is due to sex-linked meiotic drive, which can also result in degeneration and loss of single-copy ancestral genes. Another theory for ampliconic Y-linked gene families is that the amplicons allow for intrachromosomal rearrangements and therefore prevent further Y-degeneration (Rozen et al., 2003). Therefore, the amplified gene families may have arisen due to genomic competition between the X and Y chromosome during spermatogenesis (Bachtrog, 2014).

1.4.1 *Sry*

Of the twelve protein-coding genes on the mouse Y chromosome short arm, the gene *Sry* is considered the "testis-determining" factor. The human *SRY* gene was discovered in 1990 (Sinclair et al., 1990) and the homologous gene *Sry* was discovered in mice shortly after (Gubbay et al., 1990). Integration of an *Sry*-containing transgene into XX female mice resulted in sex reversal (Koopman et al., 1991).

Sry is a single-exon gene and member of the SOX (Sry-related high motility group (HMG) box) family of transcription factors (Gubbay et al., 1990). The SOX family have a variety of functions during embryogenesis and in differentiation (Kamachi and Kondoh, 2013). The HMG box controls the essential function of the transcription factors by binding DNA and inducing a bend (Harley and Goodfellow, 1994, Harley et al., 1994). Many XY female patients have mutations in *Sry* encoding for the HMG domain, highlighting the essential function of this region. Human and goat *SRY*/*Sry* transgenes that have been inserted into XX mice function equivalently and the mice develop as XX males (Pannetier et al., 2006, Lovell-Badge et al., 2002). Although the HMG domain is highly conserved amongst species, the remaining sequence conservation of *Sry* is extremely poor (Kashimada and Koopman, 2010).

Sry expression in the mouse is extremely tightly regulated within the gonad. During embryogenesis, *Sry* expression is first detectable by E10.5 in the genital ridge somatic cells. Expression peaks at E11.5 and is lost by E12.5 (Koopman et al., 1990, Kashimada and Koopman, 2010, Jeske et al., 1995). In humans, *SRY* is more widely expressed. Quickly after *Sry* is upregulated in the mouse gonad, a second gene, *Sox9* (SRY-Box 9) is also upregulated. Unlike *Sry* expression which is only active for a short period of time, *Sox9* expression is maintained beyond E12.5. *Sox9* knock-out in XY mice (Chaboissier et al., 2004, Barrionuevo et al., 2006), and *Sox9* overexpression in XX mice (Bishop et al., 2000, Vidal et al., 2001), both show sex-reversal phenotypes. These studies have implicated upregulation of *Sox9* in male sex-determination. Sekido and Lovell-Badge discovered the "testis-specific enhancer of *Sox9* core" (TESCO), a 1.4kb region upstream of the *Sox9* transcriptional start site. They showed that SRY with SF1 (Steroidogenic factor 1) bind directly to TESCO and induce upregulation of *Sox9*, highlighting evidence for *Sry* being a transcriptional activator (Sekido and Lovell-Badge, 2008). Deletion of TESCO in mice results in a reduction of *Sox9* expression (Gonen et al., 2017). Deletion of a *Sox9* upstream regulatory element called Enhancer 13 results in XY females (Gonen et al., 2018).

1.4.2 *Eif2s3y*

The hunt for the 'testis-determining factor' resulted in the discovery of the Y chromosome linked gene *Sry*. XX females containing *Sry* transgenes are sex-reversed and show male phenotypic characteristics (Koopman et al., 1991, Koopman et al., 1990). However the XX sex-reversed mice are infertile because they do not produce mature sperm. A study in 2001 by Mazeyrat *et al* showed that the essential Y-linked factor for mouse spermatogenesis is a subunit of the gene *Eif2*, called *Eif2s3y.* They utilised a mouse line containing a Y chromosome short arm partial deletion. Males carrying the deletion do not undergo spermatogenesis. They systematically introduced a number of Y-linked genes as transgenes. Upon introduction of *Eif2s3y*, spermatogenesis was restored (Mazeyrat et al., 2001). In 2009, studies by Yamauchi *et al* showed that male mice that lack the entire long arm of the Y chromosome can still produce offspring when sperm are injected into oocytes via "intracytoplasmic sperm injection" (ICSI), suggesting that genes on the Yp

are responsible for spermatogenic function. XO mice that contain an *Sry* gene and an *Eif2s3y* transgene are able to produce round-spermatids. Yamauchi *et al* (2014) showed that upon injection of *Eif2s3y* transgenic XO*Sry* round spermatids into oocytes by "round spermatid oocyte injection" (ROSI), the oocytes were successfully fertilised. Successfully cleaved embryos were transferred into pseudopregnant recipient females and live offspring were obtained (Yamauchi et al., 2014). These studies show that Y chromosome contribution of just two genes, *Sry* and *Eif2s3y,* is sufficient to generate live offspring by assisted reproductive technologies.

1.4.3 *Uty*

The role of *Sry* and *Eif2s3y* in male-specific development has been clearly defined. However for some Y chromosome-linked genes the role in development is more uncertain. An example of this is *Uty*. The Y chromosome linked *Uty* transcript is 5.5kb (Greenfield et al., 1998, Greenfield et al., 1996) and encodes a histone demethylase that demethylates the trimethylation on H3K27me3 (Walport et al., 2014). In humans, *Uty* is expressed in multiple tissue types, including the spleen and thymus, and male-specific tissue such as the prostate and testis (Yang et al., 2018). Studies have suggested that *Uty* may be involved in a gene regulatory network underlying development of prostate tissue and that dysregulation of *Uty* may predispose men to prostate cancer (Dutta et al., 2016).

The role of the X and Y chromosome genes in female and male development, and the impact of the sex chromosomes in factors such as susceptibility to disease, is an ever increasing field of research. The advent of CRISPR-Cas9 genome editing has provided an ease in which to investigate even more deeply the role of sex chromosome-linked genes, generate targeted X- and Y-linked transgenics, and begin to tackle previously unanswered questions regarding the role of the sex chromosomes outside the gonad. However the high complexity of the Y chromosome continues to provide a challenging environment in which to perform genome editing. This will be discussed in further detail in 6.1.

1.5 Development of CRISPR-Cas9 as a genome editing tool

"Clustered regularly interspersed short palindromic repeat" (CRISPR) based genome editing has revolutionised diverse fields of biological research. Consisting of two components; a Cas9 endonuclease and single guide RNA (sgRNA); the technology is quickly and easily manipulated to target many genes or genomic regions. Furthermore, the range of techniques in which CRISPR-Cas9 genome editing is now utilised has expanded far beyond simple knock-out insertion/deletion mutations. The expansion of the CRISPR-Cas9 toolbox now includes variants such as catalytically "dead" or "active" Cas9 to fine-tune expression of genes, and/or regulate chromatin and epigenetics.

1.5.1 A brief history of CRISPR-Cas

CRISPRs were originally discovered in 1987 in *E. coli*. "CRISPRs" refer to repetitive elements interspersed with non-repetitive sequences called spacers (Ishino et al., 1987). Although discovered in the genome of many bacterial and archaeal species, the non-repetitive spacers were homologous to viral and mobile genetic element DNA (Bolotin et al., 2005, Mojica et al., 2005, Pourcel et al., 2005). It was found that Cas (CRISPR-associated) genes are very closely linked to CRISPR elements, are highly conserved, and encode putative nuclease and helicase domains (Doudna and Charpentier, 2014, Bolotin et al., 2005, Haft et al., 2005, Jansen et al., 2002, Pourcel et al., 2005). Many hypothesised that the Cas genes are expressed upon phage invasion and target the invading viral DNA for DNA cleavage (Bolotin et al., 2005, Pourcel et al., 2005, Mojica et al., 2005). The integrated spacer sequence from a previous viral infection of the bacteria guides the Cas endonuclease to the newly invading virus (Doudna and Charpentier, 2014, Makarova et al., 2006, Barrangou et al., 2007, Marraffini and Sontheimer, 2008). However it was not until 2011 that the precise mechanism of Cas-regulated adaptive immunity was discovered.

Upon invasion by the virus or phage, the host bacteria expresses a non-coding pre-CRISPR RNA (pre-crRNA) which is processed to produce the mature CRISPR RNA (crRNA) from the spacer sequence. The mature crRNA molecule consists of a region that

is complementary to the viral DNA, and a repeat rich-region. RNA sequencing of *Streptococcus pyogenes* revealed the expression of a second RNA, called the trans-encoded small RNAs (tracrRNA; Deltcheva et al., 2011). These small tracrRNAs contain complementarity to the repeat rich region of the crRNA. The tracrRNA also has a vital role in the maturation of the pre-crRNA to the mature crRNA by the activity of host factors RNase III and Csn1 protein, another CRISPR-associated protein (Deltcheva et al., 2011). Doudna and colleagues discovered that all phage spacer sequences inserted into the bacterial host genomes also contained a similar sequence motif, called the protospacer adjacent motif (PAM; Jinek et al., 2012, Doudna and Charpentier, 2014). The precise sequence of the motif varied in different bacterial species with different Cas genes, but was consistent within species. The PAM sequence in the viral sequence is recognised by the Cas nucleases. In 2012, Jinek *et al* showed that the crRNA and tracrRNA could be produced as one chimeric single guide RNA ("sgRNA") by fusion of the two RNAs. This meant that a sgRNA could be synthetically produced to target virtually any region of the genome, provided a PAM sequence was present at the target site. A schematic of discovery and development of CRISPR-Cas technology to become a genome editing tool is described in more detail in 1.5.2 and summarised in Figure 4.

Thousands of studies have used CRISPR-Cas9 genome editing technology and the majority of these have utilised a Cas gene derived from *S. pyogenes,* called SpCas9 . *S. pyogenes* Cas9 is from the type II CRISPR system (contained with class II systems), commonly occurring in bacteria (class II also contains type IV, V and VI). Class I systems (type I and III) are more commonly used by Archaea. The use of the *S. pyogenes* CRISPR-Cas9 system is predominantly because of the simplicity of the sgRNA and single Cas9 protein (Doudna and Charpentier, 2014, Jinek et al., 2012, Gasiunas et al., 2012), and the ease of design; the recognised PAM is a trinucleotide NGG in the 5' to 3' direction, providing a large number of target possibilities in the mammalian genome. The crRNA component of the sgRNA is designed to be complementary to the target sequence (usually 20 nucleotides long) upstream of the PAM site, generating a simple "two component" system (Doudna and Charpentier, 2014, Jinek et al., 2012).

However there is also a growing interest in using other CRISPR-Cas systems, such as Cpf1 proteins (e.g. from *Acidaminococcus* and *Lachnospiraceae*) which generate a staggered end DNA double strand break (DSB), compared to the blunt end generated by SpCas9. There is also a growing number of Cas9 enzymes harnessed from other bacterial systems, such as *Staphylococcus aureus* (SaCas9), and *Campylobacter jejuni* (CjCas9). These species have smaller Cas9 gene sequences, making them easier to package into different viral vectors. The main disadvantage of these newer systems is the higher complexity of the required PAM sequence (5' to 3'; SaCas9: NNGRRT, and CjCas9: NNNNACAC), reducing the number of available targets in the genome.

Irrespective of which CRISPR-Cas system is harnessed to generate DNA DSBs, the endogenous cell machinery repairs the DNA break in a manner similar to when DNA DSBs occur endogenously. Dependent largely on the stage of the cell cycle, the DNA DSB is usually repaired in one of two ways. The first possibility of repair mode is by non-homologous end joining (NHEJ; Figure 3a), the second is by homology directed repair (HDR; Figure 3b). The mechanism of these two main methods of DNA repair is described in more detail below.

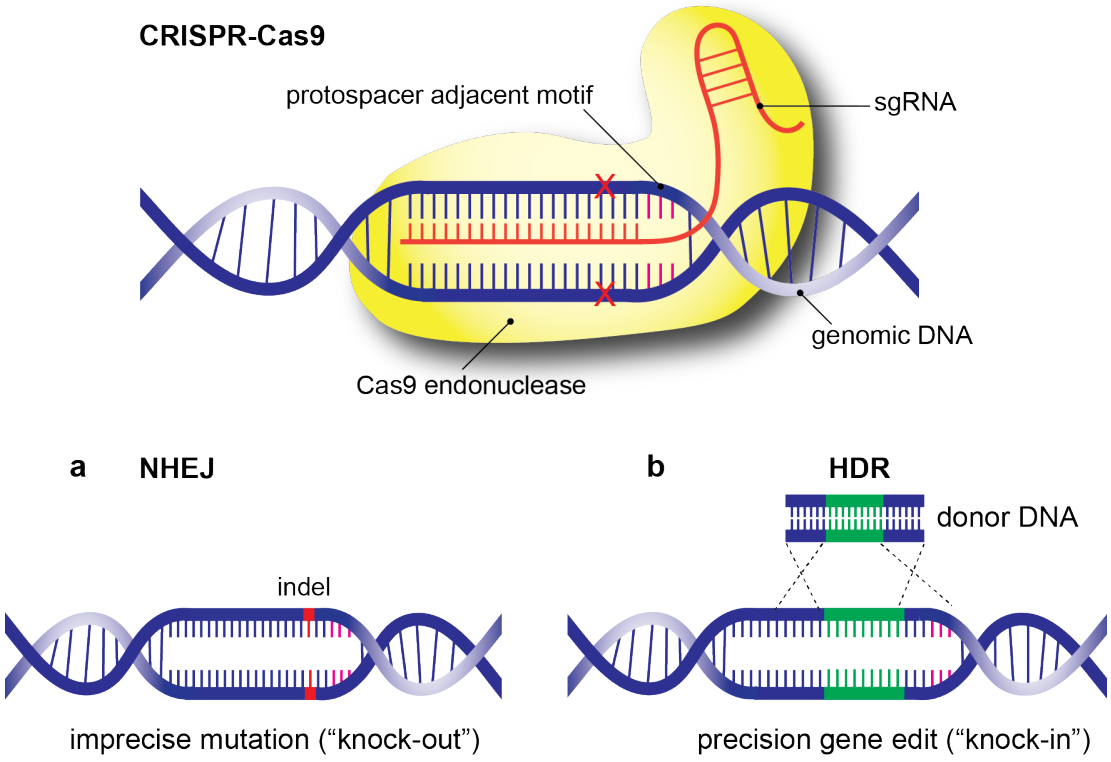

Figure 3. Schematic of CRISPR-Cas9 genome editing modes of producing mutations

CRISPR-Cas9 induced mutations, guided by a sgRNA to a 20 nucleotide target region immediately adjacent (5’) to a protospacer adjacent motif (PAM) trinucleotide NGG. The Cas9 endonuclease cuts at the -4 nucleotide position (red X) inducing a blunt end double stranded DNA break in the genomic DNA. The cleavage is repaired by one of two main mechanisms (a) non-homologous end joining (NHEJ) to generally produce a knock-out mutation, or (b) homology directed repair (HDR) to generate a knock-in.

1.5.1.1 Non-Homologous End Joining

DNA DSBs are repaired using endogenous cellular DNA-repair pathways. DNA DSBs induced by introduction of exogenous CRISPR-Cas9 components are repaired using the same DNA DSB-repair mechanisms. One mechanism of DNA repair is by non-homologous end-joining (NHEJ). NHEJ is the predominant form of DNA repair because there is no requirement for the presence of a repair template, such as a sister chromatid. Due to this lack of requirement for any kind of repair template, the mode of NHEJ is not limited to any particular stage of the cell cycle (Lieber, 2010, Maruyama et al., 2015). The mechanism of repair by NHEJ can be broken down into four steps: 1) recognition of the DNA DSB and assembly of the NHEJ complex at the break site. 2) bridging the DNA

DSB and stabilising the broken ends. 3) further processing of the break site. 4) ligation of the DNA strands. These steps are described in further detail below.

In order to repair the DNA DSB by NHEJ, firstly the broken ends are recognised by proteins Ku70 and Ku80 (also known as Ku86/*Xrcc5;* X-ray repair complementing defective repair in Chinese hamster cells 5) which form a heterodimer. This protein complex forms a ring-like structure which slides onto the DNA at the DSB, thereby binding the backbone of the DNA (Walker et al., 2001). This protein complex then acts as a scaffold to recruit further proteins to form the NHEJ complex. These factors include DNA-PKcs (DNA-dependent protein kinase), XRCC4 (X-ray repair cross complementing 4), DNA ligase IV, XLF (XRCC4-like factor) and APLF (Aprataxin and PNKP like factor; Davis and Chen, 2013). Ku70/Ku80 recruits DNA-PKcs to form an active complex in the presence of DNA (Gottlieb and Jackson, 1993), inducing DNA-PKc kinase activity. XRCC4 directly interacts with the Ku70 subunit (Mari et al., 2006), whilst DNA ligase IV interacts with the heterodimer (Costantini et al., 2007, Hsu et al., 2002). Further processing enzymes are then recruited to the DNA DSB. More recently, studies suggested that APLF may function to stabilise the whole NHEJ complex (Grundy et al., 2013, Rulten et al., 2011).

The NHEJ complex holds the DNA DSB ends in close proximity, forming a paired-end complex that stabilises the DSB. Essentially, it is the recruitment of DNA-PKcs by the Ku70/Ku80 heterodimer that enables the formation of the complex holding the two DSB ends in a stable formation (Cary et al., 1998, Weterings and van Gent, 2004). Furthermore, the XRCC4-XLF components may form a filamentous structure that further plays a role in bridging the broken ends (Malivert et al., 2010, Callebaut et al., 2006). To process the DSB site further, additional processing enzymes are recruited by the Ku70/Ku80, including Artemis, PNKP (Polynucleotide kinase 3-phosphatase), APLF, Polymerases μ and λ, WRN (Werner syndrome RecQ like helicase), Aprataxin and Ku (reviewed in (Davis and Chen, 2013). For example, Artemis, possibly activated by phosphorylation, is known to have nucleolytic activity by nicking 5' overhangs (Li et al., 2014a, Chang et al., 2015). Polymerase μ fills in DNA gaps by polymerising from a template strand using dNTPs and rNTPs (Nick McElhinny and Ramsden, 2003).

The paired end complex then ligates the DNA DSB ends together, repairing the DNA break. XRCC4 stabilises DNA ligase IV which induces the activation of DNA ligase IV by adenylation (Grawunder et al., 1997). The adenylated active DNA ligase IV ligates the processed ends across the DSB site, to form the repaired dsDNA helix (Grawunder et al., 1997, Gu et al., 2007a, Gu et al., 2007b).

The NHEJ repair mechanism is essential for intentional introduction of mismatches at the DNA DSB site when using CRISPR-Cas9 for genome editing. The continual DNA DSBs at target sites caused by CRISPR-Cas9 requires a constant NHEJ repair. Due to the lack of repair template, NHEJ is therefore error-prone. Eventually, the NHEJ will introduce mismatches or errors at the target site. These errors could be introduction of nucleotides ("insertion") or loss of some nucleotides ("deletion"), together known as "indels". The mutation could also be replacement of a single nucleotide (single nucleotide variant, SNV), resulting in a mutation at that target site. There is a two-thirds chance of generating a frame-shift mutation by NHEJ due to disruption of a trinucleotide codon. Frame-shift mutations are more likely to produce loss-of-function phenotypes when occurring in the reading-frame of a protein-coding gene. Introduction of indel mutations in the target DNA will prevent further CRISPR-Cas induced DSBs, as the sequence recognised CRISPR-Cas components is no longer intact. The use of CRISPR-Cas9 tools to generate loss-of-function mutants in a variety of mammalian cell types is described further below (see 1.5.2).

1.5.1.2 Homology Directed Repair

Unlike NHEJ, which does not require a DNA template, homology directed repair (HDR) requires a repair template. Due to the requirement for a repair template, usually the sister chromatid, HDR is limited to the S and G2 phases of the cell cycle (Maruyama et al., 2015, Zhao et al., 2017, Escribano-Diaz et al., 2013). Therefore, HDR is a significantly less common mode of DNA DSB repair (Chu et al., 2015, Maruyama et al., 2015).

The model of DNA DSB repair by homologous recombination (HR) was first postulated in 1983 (Szostak et al., 1983) with a description of DNA DSB repair via a double Holliday junction, based on the Holliday junction model by Robin Holliday (Holliday, 2007). Current understanding of HR proposes there are several steps to induce DNA DSB repair (reviewed in (Jasin and Rothstein, 2013). Initially, the DSB is resected at the 5' end to generate a 3' overhang (Kass and Jasin, 2010, Symington and Gautier, 2011). The DSB ends are processed by a complex of Mre11/Rad50/Xrs2 (MRE11 homolog, double strand break repair nuclease/RAD50 double strand break repair protein/X-ray sensitive), termed the MRX complex in *Saccharomyces cerevisiae* (Usui et al., 1998), or Mre11/Rad50/Nbs1 (MRE11 homolog, double strand break repair nuclease/RAD50 double strand break repair protein/Nibrin), the MRN complex, in mammals (Dolganov et al., 1996, Trujillo et al., 1998). In mammals, CtIP (CtBP-interacting protein, homologous to Sae2 in yeast) is recruited to DSBs in G2/S phase and recruits RPA (Replication protein A) and ATR (Ataxia telangiectasia and Rad3 related) to the DSBs (Sartori et al., 2007). If the 5' overhang is extensive, there is further recruitment of ExoI (Exonucolease 1), or a complex of Sgs1/Dna2 (Mimitou and Symington, 2008, Zhu et al., 2008). Completion of strand resection and production of the 3' overhang allows for strand invasion and repair by DNA synthesis. The invading DNA strand then displaces one strand of the double stranded DNA duplex and pairs with the remaining single stranded DNA, resulting in the formation of 'hybrid' DNA known as a displacement loop ("D loop"). The recombination intermediate is then resolved to complete the DNA repair process. The Ku complex, which has an essential role in NHEJ, may block DSB-end resection and inhibit HR. This is supported by evidence that Ku knock-out increases the efficiency of HR (Pierce et al., 2001).

HDR of DNA DSBs has been utilised experimentally to insert exogenous DNA sequences into the genome. Knock-in by HDR was originally shown in studies that utilised endonucleases such as I-SceI with a known DNA cleavage sequence. A I-SceI expression vector was introduced into cells contemporaneously with a plasmid containing repeats of homologous sequence. DNA DSBs were introduced in the target chromosome, and not in the plasmid (Rouet et al., 1994, Jasin et al., 1985). The genomic DSB was repaired

either by the plasmid, by the sister chromatid or by error-free ligation of the cleavage site (Rouet et al., 1994).

The strategy of HDR to generate a 'knock-in', such as LoxP sites or fluorescent reporters has been developed to produce many genome engineered laboratory animal models. The sequence of interest is flanked by large lengths of sequence homologous to the target region in the genome. With the advent, in 2013, of CRISPR-Cas9 for genome editing in mammalian cells (Cong et al., 2013, Mali et al., 2013, Jinek et al., 2013), it is now relatively easy to generate knock-in transgenes. In this strategy, the CRISPR-Cas9 components targeting a gene of interest can be introduced simultaneously with a repair template containing homology regions flanking the transgene ("homology arms"). In this way, the gene is targeted for DNA DSBs, whilst also providing a template to repair the DNA break. HDR is also being used to generate targeted single nucleotide changes in a very precise manner. Applications, examples and developments of generating transgenics by HDR are described further in 1.5.2.

1.5.2 Utilising CRISPR-Cas *in vitro* and *in vivo* to generate transgenic lines

CRISPR-Cas9 tools are now in place to generate knock-out by NHEJ and knock-in transgenics by HDR at a target locus with relative speed and ease, as well as more complex regulation of gene expression. To generate indel mutations for loss-of-function studies, the Cas9 and sgRNA are introduced *in vitro* or *in vivo*. For HDR experiments, the Cas9 and sgRNA targeting the locus of interest are contemporaneously introduced *in vitro* or *in vivo* with a repair template. Use of CRISPR-Cas9 for gene editing began in 2013 (Jinek et al., 2013, Cong et al., 2013, Wang et al., 2013b, Mali et al., 2013, Yang et al., 2013a), and is being increasingly widely-adopted, with design and strategy becoming ever more efficient and optimised.

The first three studies in 2013 that harnessed CRISPR-Cas9 genome editing did so in mammalian cells, highlighting a breakthrough in transferring a bacterial immunological response system into a mammalian context. Jinek *et al* transfected human HEK293T cells with human codon-optimised SpCas9 fused to eGFP for visualisation of transfection, and

an sgRNA targeting a gene *CLTA* (Clathrin light chain A). The results of this study showed a 6-8% mutation efficiency at *CLTA*. Jinek *et al* suggested that the efficiency of indel mutation could be improved by complexing sgRNA and Cas9 prior to transfection, to increase the stability of the sgRNA (Jinek et al., 2013). Cong *et al* also used human codon-optimised SpCas9, fused to multiple nuclear localisation signals (NLS) to increase the efficiency of Cas9 localisation to the nucleus. The SpCas9 was co-transfected in human 293FT cells with three further factors: separate tracrRNA and pre-crRNA, and RNaseIII, targeting the gene *EMX1* (Empty spiracles homeobox 1). RNaseIII was described in a study by Deltcheva *et al* (2011) as an essential component required for the maturation of pre-crRNA to crRNA (Deltcheva et al., 2011). Cong *et al* showed that addition or removal of RNase III in the transfection mixture had no noticeable effect on mutation efficiency, suggesting that mammalian cells can use endogenous RNases for maturation of the pre-crRNA into crRNA. Conversely, removal of either the tracrRNA or pre-crRNA resulted in no indel mutations induced at EMX1. This study highlighted a minimal three-component capacity for generating mutations at a target locus (Cong et al., 2013). Cong *et al* also showed that mutations could be generated in a target gene at a higher efficiency by transfection of a single plasmid that expresses both the Cas9 and sgRNA, instead of as separate components, improving the simplicity of CRISPR-Cas9 based genome editing (Cong et al., 2013). Mali *et al* (2013) was the first example of generating a knock-in by HDR. They utilised a human HEK293T cell line that contains a eGFP reporter. Expression of the eGFP reporter is inhibited due to the presence of a stop codon at the beginning of the coding sequence. Mali *et al* transfected the HEK293T cells with an sgRNA and Cas9 targeting eGFP, alongside a repair template that removes the stop codon. Integration of the repair template by HDR thereby induces eGFP expression and results showed that the eGFP positive cells were successfully selected by fluorescence activated cell sorting (Mali et al., 2013).

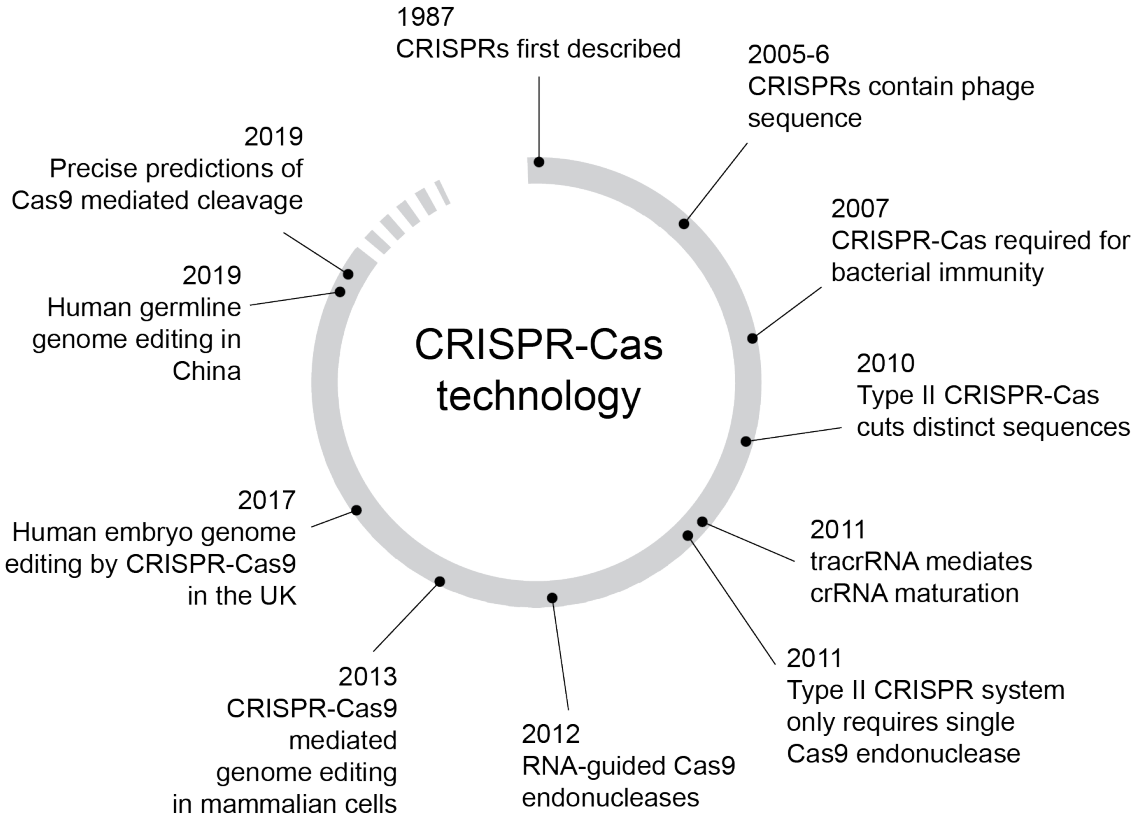

Figure 4. Timeline of CRISPR-Cas technology development

A timeline highlighting key discoveries in the development of CRISPR-Cas9 as a technology for genome editing. Originally discovered as a immunological defence mechanism against invading phage, the action of Cas endonucleases targeting specific sequences, guided by chimeric sgRNAs has been harnessed to target virtually any region of the genome in mammalian cells *in vitro*, and *in vivo*.

Further studies in 2013 expanded the toolbox of CRISPR-Cas9 genome editing by development of tools that allowed for targeting of multiple genes at once, called "multiplexing" (Ran et al., 2013a, Wang et al., 2013b). Wang *et al* introduced five different sgRNAs (plus Cas9) targeting the genes *Tet1*, *Tet2*, *Tet3* (Tet methylcytosine dioxygenase 1, 2, 3), *Sry* and *Uty* into mouse ESCs, and mutations were generated in all targeted genes (Wang et al., 2013b). A single plasmid vector was generated by Ran *et al* which contained two sgRNA scaffolds for expression of two sgRNAs from a single vector. This could be used for multiplexing of different genes, or for two sgRNAs targeting the same gene, resulting in large deletions of the intervening sequence (Ran et al., 2013a). Large deletions of the target gene has the potential to increase the likelihood of a loss-of-function mutation.

In many of these early studies, the aim was to generate a mutation in a target gene by CRISPR-Cas9 induced DNA DSBs. The DNA DSB is repaired by NHEJ and is error-prone, producing an indel mutation. Recently, evidence has shown that the mutational outcome at a specific target site can be predictable (see 1.5.3; Chakrabarti et al., 2019, van Overbeek et al., 2016, Allen et al., 2018). Nevertheless, repair by NHEJ carries a risk of variability, large indels, and/or complex rearrangements (Kosicki et al., 2018) which can make genotyping challenging. To increase the accuracy of assessing genetic modifications, researchers are utilising the HDR mode of DNA DSB repair, to either make small precise nucleotide changes, or to introduce exogenous DNA. The sequence of interest to be integrated into the target genome is flanked by homology arms, which are homologous to the target genome. This sequence of interest could be very small, for example a single nucleotide to replace the existing nucleotide in the target gene. The sequence could also be large, for example LoxP sites to generate conditional knock-outs, or fluorescent reporter genes for gene-tagging. The repair template (or "targeting vector") is introduced *in vitro* or *in vivo* alongside Cas9 and an sgRNA targeting the gene of interest.

Targeting experiments have traditionally used dsDNA plasmids as the repair template to insert transgenes. Plasmids are often large, supercoiled DNA and can be challenging for packaging the DNA into cells for HDR. The early study by Wang *et al* (2013) showcased a different type of repair template for introducing very specific point mutations at a target locus. They co-injected target sgRNAs, Cas9 mRNA and single stranded oligonucleotides (ssODNs) to generate a knock-in (Wang et al., 2013b). The use of ssODNs is gaining traction for inserting small transgenes, for example point mutations, as the DNA is smaller and easier to package into cells (e.g. Wang et al., 2013b). Quadros *et al* (2017) also performed knock-in experiments using long single-stranded oligonucleotide (lssODN) repair templates, termed "Easi-CRISPR". The efficiency of knock-in with lssODNS was extremely effective, with efficiencies of 100% (Quadros et al., 2017). The ssODNs/lssODNs can be manufactured either in-house by molecular biology techniques, or synthetically, providing a quick and easy repair template for small knock-in experiments. However lssODNs generated synthetically are restricted to the maximum size private companies can manufacture with high fidelity.

HDR can also be utilised to knock-in large transgenes into a target gene, for example to generate fluorescent reporter 'tagged' proteins. Gene-tagging can be useful for visualising the expression domains of a protein-coding gene. Yang *et al*, (2013) highlighted examples of generating fluorescent reporter knock-ins for tagging a number of essential embryonic pre-implantation genes, marking different lineages. In this study, reporter alleles for genes *Nanog* (mCherry), *Sox2* (SRY-box 2; V5) and *Oct4* (eGFP) were generated using a 'one-step' knock-in system in mice (Yang et al., 2013a). For large insertions such as fluorescent reporters (e.g. Yang et al. 2013), or replacement of genes/exons with antibiotic selection cassettes (e.g. Spiegel et al., 2019), CRISPR-Cas9 HDR targeting strategies usually make use of dsDNA plasmid repair vectors. The targeting vector length with long homology arms is usually greater than that possibly produced by lssODN synthesis.

The design of the homology arms (HAs) in gene targeting experiments is an essential consideration. A study by Richardson *et al* in 2016 aimed to produce a trinucleotide change in a target gene by CRISPR-Cas9 HDR. They used lssODN repair templates with varying lengths of 5' and 3' HAs, complementary to either the target or non-target strand. They observed that repair templates with homology to the non-target strand produced a successful knock-in at a greater efficiency than at on-target strands by HDR (Richardson et al., 2016) also observed in other studies (Lin et al., 2014, Yang et al., 2013b). Asymmetric HAs produced the highest rate of integration by HDR compared to symmetric HAs, particularly when the longer HA is on the PAM-proximal side of the DSB (Richardson et al., 2016). Furthermore, the closer the HAs align to the DNA DSB position, the higher rate of repair template insertion (Cong et al., 2013, Mali et al., 2013, Paquet et al., 2016, Kwart et al., 2017). In order to have homology arms as close as possible to the DSB point, the target sgRNA sequence may also be contained within the homology arm. Having the sgRNA sequence contained within the homology arm is disadvantageous for two reasons. Firstly; the CRISPR-Cas9 system may induce DNA DSBs in the repair template, inhibiting accurate HDR. Secondly, the repair template may be successfully integrated by HDR, but lost again due to CRISPR-Cas9 targeting of the HDR allele. To circumvent issues with CRISPR-Cas9-induced DNA DSBs in the repair

template, a silent mutation can be introduced into the PAM sequence in the homology arm-containing sgRNA sequence, called a 'blocking mutation' (Paquet et al., 2016). Despite considerations of homology arm length and proximity to the sgRNA target site, the efficiency of HDR is poor compared to the NHEJ (Chu et al., 2015, Maruyama et al., 2015). Addition of chemicals that inhibit NHEJ, such as SCR7 which functions by inhibiting DNA ligase IV (Ma et al., 2016, Maruyama et al., 2015, Srivastava et al., 2012), or that increase the rate of HDR, such as RS-1, which enhances RAD51 DNA binding (Jayathilaka et al., 2008, Song et al., 2016, Pinder et al., 2015), have been shown to improve HDR rates in some cell types but they have not widely employed in many species.

1.5.3 Predicting CRISPR-Cas9 mutational outcomes

Since the adaption of CRISPR-Cas for genome editing, researchers are now investigating further the structure and endonuclease action of Cas enzymes, in particular SpCas9. Structural studies of Cas9 have been essential in the understanding of how and where DNA DSBs are induced at the target site, and how the location of the DSB affects the mutational outcome at the target locus.

SpCas9 is bilobed, containing an α-helical recognition (REC) lobe, and a nuclease (NUC) lobe, bridged by an α-helix (Jinek et al., 2014, Nishimasu et al., 2014, Sampson and Weiss, 2013). The NUC lobe has two major nuclease domains; the RuvC and HNH domains. The DNA DSB endonuclease function of both the RuvC and HNH domains is Mg2+ dependent (Jinek et al., 2012, Gong et al., 2018). The HNH domain induces DNA cleavage at the DNA strand that is complementary to the 20 nucleotide crRNA sequence, whilst the RuvC domain cleaves the opposite DNA strand (Jinek et al., 2012, Gasiunas et al., 2012). The bilobed Cas9 endonuclease induces a blunt DSB at a specific point, between nucleotides 3 and 4 in the 5' direction ("-4") to the PAM site (red X, Figure 3; Shen et al., 2013). A mutation in either the HNH or RuvC domain results in a DNA single strand break (SSB). These Cas9 variants are nickases ("Cas9n"). A mutation in both RuvC and HNH domains results in complete loss of the endonuclease cleavage activity (Jinek et al., 2012, Gasiunas et al., 2012).

The discovery that Cas9 endonuclease induces a blunt DNA DSB at the -4 position opened up research avenues into predictions of precise CRISPR-Cas9 induced mutational outcome. In 2016, a study by van Overbeek *et al* used 96 sgRNA sequences for different targets and assessed the indels at each target (van Overbeek et al., 2016). Although the number of sgRNA targets was small, this study was the first to suggest that mutational outcome was non-random. In 2018, a larger-scale study by Allen *et al* utilised over 40,000 synthetic constructs to assess mutation outcome. Each construct contained both the sgRNA expression cassette, and the sgRNA target sequence. The constructs were introduced into Cas9-expressing cell lines and the target loci sequenced. This study showed there is increased probability of the mutational outcome at a certain sites being predictable, based on the surrounding sequence (Allen et al., 2018). Chakrabarti *et al* (2019) improved on these studies with precise predictions of specific nucleotide indels. They showed that there is much higher likelihood of the mutation at the target site being a mononucleotide insertion if the -4 position is an A or T nucleotide. If the -4 position contains a dinucleotide of two C nucleotides, the most commonly predicted outcome is a mononucleotide deletion of one of the C nucleotides. Lastly, if the -4 position is a G nucleotide, the outcome is much more unpredictable, with varying types of indels at the target site (Chakrabarti et al., 2019). These results were similarly presented in a study by Shen *et al* (2018), who highlighted that in 5-11% of sgRNA target sites in the human genome, the repair outcome is predictable (Shen et al., 2018). These significant results highlight the importance in strategizing sgRNA design and target site location.

1.6 The requirement for specific sexes in the modern world – health, agriculture and pest control

Animals and animal products have been selected for use in many aspects of modern day living. This includes in health applications and life sciences; for example development of medicines such as antibiotics and other essential drugs. Farm animals have been through generations of breeding for desired genetic traits. This has been for a variety of different qualities, from reproductive ability, to growth efficiency and food production. This process of ‘artificial selection’ or ‘selective breeding’ has ensured the most productive

genetic traits are carried forward to the next generation. The male in the 'herd' therefore contributes highly to the genetic contribution of the population on average, as generally the agricultural population is made up of very few, possibly only one, male, and many more females. Often a surplus of female agricultural animals are required compared to their male siblings. This requirement for females has resulted in widespread culling of the unrequired males. Examples of these are in the layer hen and dairy cow industries, where there has been extensive development of methods to generate single-sex litters, in an attempt to reduce the need for post-natal culling of the unrequired sex.

Another positive reason to generate single-sex litters is that of adhering to the United Kingdom Home Office "3Rs", for reduction, replacement and refinement. This policy aims to reduce unnecessary animal culling in laboratories across the UK. Many research groups studying sex-specific biology will also produce many animals of the unrequired sex, which are usually culled post-natally. A genetic method for reducing the number of the unrequired sex being born would have positive implications for reducing this animal culling issue.

1.6.1 Layer hens

In the layer hen industry, egg production for food is entirely from females, and the male chicks are culled shortly after birth. Estimates of exactly how many male chicks are culled per year vary, but are thought to be between 6-7 billion (Krautwald-Junghanns et al., 2018). This is an extreme example of where the requirement for females outweighs the requirement for males, and is generating a vast animal welfare issue with charities such as the RSPCA (Royal Society for the Prevention of Cruelty to Animals) and PETA (People for the Ethical Treatment of Animals).

Over the years, a number of non-invasive methods have been developed to reduce the extent of post-natal male chick culling. Hyperspectral imaging can be used on mid-incubation eggs where the males and females have different feather colours (Gohler et al., 2017), whilst morphometric quantification such as egg shape (Yİlmaz-Dİkmen and Dİkmen, 2013, Krautwald-Junghanns et al., 2018, Galli et al., 2018) or egg odour

(Webster et al., 2015) which is different between males and females, has also been described. However these techniques have not been widely employed.

Invasive methods to determine the sex of the developing chick embryo are generally more accurate, and more widely utilised as they are based on detection of female- or male-specific molecules (Clinton et al., 2016). The different sexes are determined by quantification of hormone levels in the extracted allantoic fluid. Weissmann *et al* (2013) showed that detection of differential levels of estrone sulfate in extracted allantoic fluid was sufficient to determine the sex of the developing embryo. Interestingly, measurements of testosterone gave significantly more ambiguous results. On average, female allantoic fluid contained levels of estrone sulfate at 0.312ng/ml, whereas male fluid was 0.110ng/ml. However the levels of estrone sulfate could only be detected at day 9, as there was not enough fluid present at day 7 for accurate detection of hormones (Weissmann et al., 2013). The chick embryo gestation period is 21 days, and they develop pain perception at approximately day 10.5 (Weissmann et al., 2013). Therefore invasive methods at day 9 are extremely close to the onset of pain perception. This carries a risk of causing distress and damage to the developing chick and may induce non-viability in severe cases (Rosenbruch, 1994, Rosenbruch, 1997). To access the allantoic fluid for extraction, a section of the shell and inner membrane has to be removed, in a process called “windowing”. Damage to the inner protective membrane reduces the health and hatching rate of the developing chick (Fineman et al., 1986, Speksnijder and Ivarie, 2000) and exposes the developing chick to potential pathogens from the outside environment before hatching (Brown *et al*, 1965).

A different method to sex developing chicks was published in 2018 (Galli et al., 2018). This method relies on physiological differences between male and female chicks, but is a significantly less invasive procedure, as the inner membrane is not damaged. The method is based on a previously published invasive procedure study by Galli *et al* that quantified molecular sex differences in embryonic blood. The original study successfully detected male/female sex differences but also induced a reduced embryonic viability rate of approximately 10% due to inner membrane damage (Galli et al., 2017, Galli et al., 2016). In the 2018 study, Galli *et al* performed optical spectroscopy at near infrared

radiation with Raman microscopy on day 3.5 chick embryos, upon windowing of the shell, but leaving the inner membrane intact (Galli et al., 2018). They quantified spectral excitation and fluorescence intensity from exposed extra-embryonic blood vessels. Results showed that female chicks have a fluorescence intensity significantly lower than males, on average. The study also highlighted that the procedure had no effect on the chick hatching rate. Furthermore, the analysis could be performed at embryonic day 3.5 embryos, a significant advantage over earlier invasive day 9 analyses. Once the chicks were born, they were phenotypically sexed, and this result was compared to the predicted sex from the embryonic analysis. Results showed that the embryonic optical spectroscopy predicted sex correctly in over 90% of chicks (Galli et al., 2018). Although the data shows significant overlap in the spectral range for male and females, the authors propose a mathematical correction to more accurately segregate males and females.

1.6.2 Dairy cows

Similarly to the layer hen industry, the dairy industry also requires many more females than males. The most extensively researched method to generate greater numbers of female offspring is by bull sperm 'sexing'. The process of sperm sexing selectively separates X- and Y-carrying sperm. The X-carrying sperm can then be used for assisted reproduction technologies (ART) such as *in vitro* fertilisation (IVF) of female dairy cows to increase the likelihood of generating XX female offspring.

The most developed method of sexing sperm is by flow cytometry. Sperm-sorting by flow cytometry was first described in 1982 by Pinkel *et al* who developed the technology in order to separate Y- and O- carrying sperm from the vole (*Microtus oregoni*; Pinkel et al., 1982). The first flow cytometer used for sexing sperm for agriculture was called the "Beltsville sperm sexing technology" (Johnson et al., 1989, Johnson and Clarke, 1988, Johnson et al., 1999). The flow cytometry-sorted sperm were used for ART and shown to be highly successful for rabbits (Johnson et al., 1989), pigs (Johnson, 1991, (Kawarasaki et al., 1998, Rath et al., 1997), cattle (Seidel, 1997), and sheep (Johnson, 1995, Catt et al., 1996). A limitation of the Beltsville flow cytometer was that the X- and Y-bearing sperm heads had to be removed from the tails. Tail removal ensured that the sperm heads would

be correctly orientated to be sorted by flow cytometry. The movement of attached tails disorientated the sperm and they could not be detected by the cytometer lasers and therefore would be incorrectly sorted. Although the tail-less sperm were immobile, the sperm heads were able to fertilise oocytes (Johnson and Pinkel, 1986, Johnson et al., 1987, Johnson and Clarke, 1988). However sperm-head oocyte injection was not immediately translatable to an agricultural context. A key development in the technology was a flow cytometer with a nozzle that applied forces to the motile sperm (with tails), orientating them correctly. The process of sperm orientation resulted in an increased sorting efficiency from 30% to 60% for sperm with tails (Rens et al., 1998, Johnson et al., 1989, Johnson et al., 1999). The development in the sorting technology allowed for mature sperm to be sorted, frozen and thawed for used in a standard IVF techniques.

To detect the sperm being sex-sorted by flow cytometry, DNA has to be stained. In early developments of the technology, this required removal of the sperm membrane in order for the stain to access the nucleus. This process killed the cells therefore they could not be used for ART (Johnson and Pinkel, 1986). The staining method was improved by Johnson *et al*, in the late 1980's. Johnson *et al* used Hoechst 33342 for sperm staining which did not require the membranes to be removed (Johnson et al., 1987, Johnson and Clarke, 1988, Johnson et al., 1989), thereby reducing damage to the sperm. Upon staining bovine sperm with Hoeschst 33342, the DNA content of the X-carrying sperm is 3.8% higher than the Y (Johnson et al., 1999) allowing for clear separation of the two sperm types by flow cytometry. The efficiency of sorting bovine sperm by flow cytometry is high, with efficiency of pure X-carrying and Y-carrying sperm at around 90% (Seidel, 2012). However, flow cytometry sorted-sperm used for cattle ART has been shown to reduce birth rates, compared to non-sorted sperm (Dejarnette et al., 2011, Frijters et al., 2009, Seidel, 2007). Furthermore, the birth rate is not improved by increasing the sperm number used for the ART procedure (Dejarnette et al., 2011). It is thought that the reduced birth rate is due to residual Hoescht 33342 dye from the sorted-sperm interfering with pre-implantation embryonic development (Garner, 2009). However, offspring that are successfully produced from sorted-sperm ART do not appear to show any developmental defects (Tubman et al., 2004, Seidel, 2007).

As well as the reduction in live births associated with sorted sperm, there are also practical and socioeconomic disadvantages associated with flow cytometry sperm-sorting. Flow cytometers are expensive to buy and maintain, and are time-consuming in sort-time and post-sort procedure. Sperm that are sex-sorted by this technique must be sorted fresh to ensure maximum accuracy of sex selection by DNA content. Cells are sorted one at a time, resulting in an extremely time-consuming process. Furthermore, post-sorted sperm must be frozen down appropriately for use downstream in ART (Seidel, 2007).

Given the time, expense and the reduction in live birth rate, there is now an increasing interest in improving non-flow cytometry based methods for sperm sex-selection. An example of a non-flow cytometry method was published in a study in 2019 by Chowdhury *et al.* The authors developed a method whereby a monoclonal antibody (termed "WholeMom") binds bovine Y-chromosome sperm epitopes, resulting in an aggregation of the Y-sperm heads. This allows the non-affected X chromosome sperm to be freely motile to fertilise the oocytes *in vivo* (Chowdhury et al., 2019). Another study published in 2019 by Umehara *et al* inhibited the motility of mouse X-carrying sperm by ligand activation of Toll-like receptor 7/8 (TLR7/8) which is expressed only in X-sperm. They separated X- and Y-carrying sperm by differences in motility. Following IVF with Y-carrying ("fast") sperm, 83% of pups born were male. IVF with X-carrying ("slow") sperm produced 81% female offspring (Umehara et al., 2019).

Developing a genetic method whereby the unrequired sex is never produced, would drastically reduce the animal welfare impact whilst reducing the costs and time associated with current sex-sorting methods.

1.6.3 Pest control

Whilst reducing the number of males produced in the layer hen or dairy cow industry is beneficial, on the converse, reducing the number of females may be beneficial for pest control. Reducing the number of breeding females could be used to control populations of invasive rodent species such as rats or mice or the malaria (*Plasmodium)*-carrying female mosquitos (*Anopheles).* The *Plasmodium* parasite is transmitted uniquely via

female *Anopheles* mosquitos. Therefore many research groups are investigating methods to generate male-bias populations. Strategies for reducing the number of breeding females in both rodents and mosquitos take advantage of natural Mendelian inheritance segregation, and aim to distort the expected sex ratios using gene drive methods, described further in 1.6.3.1.

1.6.3.1 Gene drive

The concept of gene drive (originally called "genetic underdominance") was first postulated independently by Alexander Sergeevich Serebrovsky and Chris Curtis. They determined that chromosomal translocations could spread quickly through a population if they caused heterozygous inferiority, by populations selecting for homozygosity (Curtis, 1968, Serebrovsky A, 1969). The notion of skewing sex ratios by non-Mendelian inheritance was then described by Bill Hamilton in 1967, who suggested that sex ratios could be distorted by a mechanism of Y-linked bias (Hamilton, 1967). With the advent and ease of CRISPR-Cas9, it is now possible to control population dynamics by *in vivo* self-driven genome editing to distort offspring sex ratios. CRISPR-Cas9 gene drives could be used as a method of pest and vector control, and species such as the mosquito, are useful models of gene drive, due to the high reproductive rate and short life span.

An early study aiming to skew mosquito offspring sex ratios was published in 2007 by Windbichler *et al*. A mosquito Y chromosome (the male determinant) was edited to carry a I-PpoI endonuclease transgene that targets the X chromosome for destruction (Windbichler et al., 2007, Burt, 2003). The endonuclease transgene was later edited to restrict I-PpoI expression to male meiosis (Galizi et al., 2014, Burt, 2003, Windbichler et al., 2007, Windbichler et al., 2008). During meiosis, I-PpoI selectively destroyed X-carrying gametes, resulting in the viability of only Y chromosome carrying sperm and vastly reducing XX offspring numbers. In 2016, a study by Galizi *et al* used CRISPR-Cas9 to target repetitive ribosomal sequences on the X chromosome, driven by a spermatogenesis specific promoter, resulting in shredding of the X and extreme male bias (Galizi et al., 2016). The non-viability effect is at the gametic level, therefore there is no effect on fertility of the population, but reduced availability of X-carrying male gametes.

One of the first examples of pre-CRISPR-Cas9 genetic modification for gene drive is using homing endonucleases, a selfish genetic element found originally in bacterial systems (reviewed in e.g.(Belfort and Roberts, 1997). This gene drive element takes advantage of endogenous DNA repair mechanisms. The homing endonuclease induces DNA DSBs, followed by insertion of the gene drive element as a repair template, into the cleaved allele by HR. This results in rapid conversion of the gene drive allele from heterozygosity to homozygosity. This strategy for mosquito population control was first described in 2011, whereby a gene for the homing endonuclease I-SceI was inserted into the mosquito genome (Windbichler et al., 2011) and the fly genome (Chan et al., 2011) and was inherited by offspring at above-Mendelian frequency.

Since 2013, CRISPR-Cas9 has increased the efficiency of synthetic gene drive systems. The Cas9 endonuclease, guided by a sgRNA, induces a DNA DSB in a target gene. Endogenous HDR uses the Cas9-sgRNA transgene as the repair template (Figure 5). Therefore the Cas9-sgRNA transgene is converted from heterozygosity to homozygosity and spreads through the population at above-Mendelian frequency. The Cas9-sgRNA transgene is designed to target female fertility or viability genes, thereby reducing the female population and suppressing population growth (Esvelt et al., 2014, Gantz and Bier, 2015). Successful CRISPR-Cas9 gene drive systems have been shown in multiple species, including *Saccharomyces cerevisiae* (DiCarlo et al., 2013), *Anopheles stephensi* (Gantz et al., 2015) and *Anopheles gambiae* (Hammond et al., 2016).

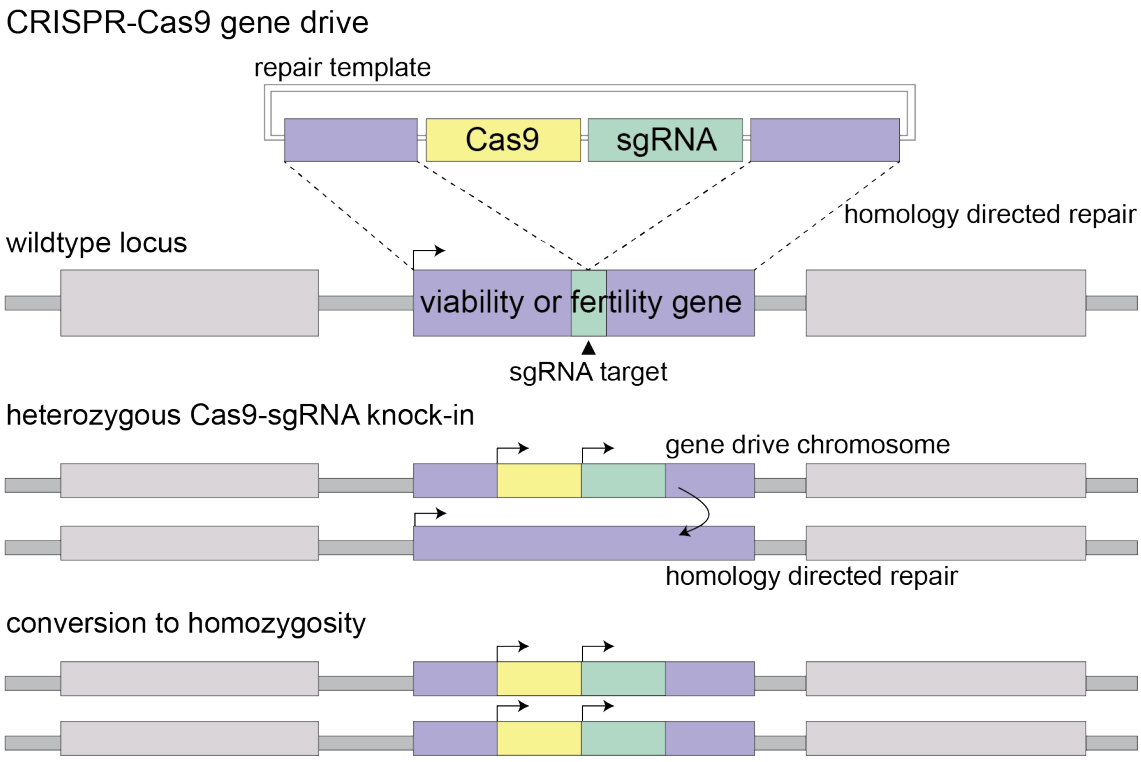

Figure 5. CRISPR-Cas9 gene drive

CRISPR-Cas9-based gene drive promotes inheritance of a Cas9-sgRNA transgene at a rate above Mendelian frequency. A dsDNA plasmid repair template contains a Cas9-sgRNA allele that targets an essential viability or fertility gene. The repair template is integrated into the target viability/fertility gene by HDR, producing a heterozygous Cas9-sgRNA genotype and a 'gene drive chromosome'. Expression of the Cas9-sgRNA transgene induces HDR at the second allele, converting the transgene from heterozygosity to homozygosity. The homozygous allele therefore spreads through the population at above Mendelian frequency.

Many of the mechanisms used to reduce pest populations has been on non-mammalian models. However in order to reduce mammalian pest populations, for example invasive rodent species, a tool needs to be developed in the mouse model. A study in 2018 by Grunwald *et al*, generated a transgenic mouse line expressing a sgRNA construct and mCherry reporter, with no Cas9 gene. The sgRNA targeted the *Tyrosinase* (*Tyr)* gene. They hypothesised that mating the sgRNA-mCherry transgenic mouse line to a Cas9-expressing mouse line would 'activate' the gene drive. Activation of the gene drive system would convert the sgRNA-mCherry allele from heterozygosity to homozygosity, named a "CopyCat element", and thereby induce a homozygous knock-out in the *Tyr* gene. Knock-out of *Tyr* results in mice with a white coat colour, and mCherry transgene expression. Furthermore, they limited Cas9 expression to the germline by using a floxed-STOP Cas9 mouse line (*Rosa26*-LSL-Cas9 (Platt et al., 2014) or *H11*-LSL-Cas9 (Chiou

et al., 2015)) and a germline promoter driving Cre. Germline-specific expression of Cas9 theoretically maximised the likelihood of HDR during germ cell mitosis/meiosis and therefore conversion of the sgRNA-mCherry allele to homozygosity. Despite Cas9 expression being limited to the germline, the study showed that the efficiency of Mendelian inheritance of the CopyCat allele was relatively poor, ranging from 0% to 72%. The majority of the offspring had mutations at the *Tyr* locus by NHEJ, without integration of the CopyCat element (Grunwald et al., 2019).

1.6.3.2 The disadvantages of gene drive

An important consideration in the mechanism of nuclease-based gene drive systems, such as CRISPR-Cas9, is the possibility of generating mutations that block progression of the gene drive system. This gene drive inhibition could be due to naturally occurring variants, appearing as single nucleotide polymorphisms (SNPs), blocking DSBs at the sgRNA recognition site. Inhibition of the gene drive mechanism could also be induced by CRISPR-Cas9 induced mutations at the target locus. The process of gene drive by CRISPR-Cas9 transgenes relies on the repair mechanism of HDR to convert the heterozygous transgene to homozygosity and therefore spread through the population at above-Mendelian frequency. However, if DNA DSBs at the target allele are repaired by NHEJ prior to HDR, indel mutations occurring at the locus would inhibit further DSBs by the gene drive system (Champer et al., 2017, Hammond et al., 2017, Hammond et al., 2016, Gantz and Bier, 2015, Gantz et al., 2015). Recent studies using mathematical modelling have suggested that the current CRISPR-Cas9 gene drive systems designed for reducing female mosquito numbers would quickly lose transmission within the population due to arising mutations at the target site and would therefore require continuous intervention (Unckless et al., 2017, Hammond et al., 2016). Furthermore, poor efficiency of gene drive in mouse models, such as that described in Grunwald *et al* (2018), means that the alleles could take a long time to spread through the population, leaving plenty of time for genetic resistance to occur.

There are a number of strategies that are currently being investigated that potentially could be implemented to prevent mutational resistance occurring in CRISPR-Cas9 gene drive (reviewed in e.g. Hammond and Galizi, 2017). One example that has gained traction

is the use of multiple sgRNAs complementary to the target gene of interest (Esvelt et al., 2014). In this strategy, if resistance occurred at the target gene by NHEJ-induced mutation, multiple other target sites are still intact for HDR of the drive allele. Another way to circumvent issues associated with CRISPR-Cas9 gene drive resistance is by harnessing an endogenous selfish genetic element found in mice called the t-haplotype. The t-haplotype is a naturally occurring gene drive element on mouse chromosome 17 (Willison and Lyon, 2000). Homozygous males are embryonic lethal, however heterozygous male mice pass on the t-haplotype element to offspring at above-Mendelian frequency (Bauer et al., 2005). Researchers are investigating the use of a t-haplotype variant which induces sterility in the males instead of non-viability, leaving females unaffected (Lyon, 2003).

1.6.4 Skewing sex ratios using a bi-component system

In 2018, Zhang *et al* highlighted a method of pest control in a lepidopteran species, using a model of *Bombyx mori* (silkworm). Silkworm species use a WZ/ZZ sex determination system and the female is the heterogametic sex (WZ chromosome complement). The authors generated a Cas9 knock-in line, linked to the W chromosome (W-Cas9). The W-Cas9 transgene is uniquely inherited by female offspring. Furthermore, a second transgenic line was generated, expressing an sgRNA targeting the *Bmtra2* (Bombyx mori transformer-2) gene. *Bmtra2* is an essential gene and is embryonic lethal when knocked-out. Mating of the W-Cas9 and sgRNA expressing lines resulted in female-specific mutations in *Bmtra2,* inducing *Bmtra2* loss-of-function and non-viability of the female progeny (Zhang et al., 2018). This study was the first example of a CRISPR-Cas9 bi-component system to induce sex-specific non-viability by knock-out of an essential gene.

In 2019, a study by Yosef *et al* used a bi-component CRISPR-Cas9 method to generate single-sex litters in mammals. They generated a Y chromosome-linked sgRNA-expressing mouse line, targeting essential genes *Atp5b* (ATP synthase subunit beta), *Casp8* (Caspase 8) and *Cdc20* (Cell division cycle 20). They theorised that mating the Y-sgRNA lines with autosomal linked-Cas9 mouse lines would result in offspring male-specific non-viability due to inheritance of both CRISPR-Cas9 components and loss-of-

function mutations in the essential genes. Inheritance of the Y-sgRNA and autosome-Cas9 transgenes significantly skewed the sex ratio of offspring, however there was not a complete loss of males. In the Y-Cas9 by Cas9 mating, 113 pups were born, 9 of which were male (8%). Six of the males did not survive through to weaning (Yosef et al., 2019). Although this study was the first study to utilise a bi-component system in the mouse model, the single-sex skew of litters was incomplete, and further improvements could be made.

In summary, there are multiple different methods associated with producing an offspring sex-skew with distinct advantages and disadvantages. Many are currently being utilised to produce single-sex litters and reduce unrequired animal culling, but they are often time-consuming and expensive. A possible cost-effective method to generating single-sex litters may be to utilise CRISPR-Cas9 components in a bi-component system. A single Cas9-expressing male in the 'herd' can be used for fertilisation (e.g. by IVF/ART) of many sgRNA-expressing females. Inheritance of both CRISPR-Cas9 components in a sex-specific manner results in embryonic non-viability of the target unrequired sex.

1.7 Aims

CRISPR-Cas9 genome editing has provided an unprecedented ease by which to generate, characterise and evaluate mutations at a target locus. Utilising the two components of the CRISPR-Cas9 technology, Cas9 and sgRNA, the aim of this thesis is to generate stable, knock-in transgenic embryonic stem cell and mouse lines that express these two components separately. Upon doing so, the "bi-component" technology can be harnessed to generate targeted mutations, utilising the unique inheritance of the male heterogametic sex chromosomes to ensure mutagenesis occurs in a sex specific manner. Inheritance of both components in an individual induces knock-out of an essential gene, via CRISPR-Cas9 mutagenesis, resulting in embryonic non-viability. This thesis makes use of the mouse laboratory model, however the technology has translational potential in agriculture, conservation and in animal welfare in conjunction with the Home Office 3Rs.

Overall, the thesis has four major results chapters:

1. In results chapter 1, I aim to investigate a "proof of principle" system with pre-existing transgenic mouse lines to ask whether it is biological possible to generate single-sex litters.
2. In results chapter 2, I aim to screen sgRNAs *in vitro* to determine efficiency of mutagenesis, characterising the dynamics and spectrum of mutations at the target locus. I then aim to generate a stable transgenic mouse line expressing a highly mutagenic sgRNA.
3. In results chapter 3, I aim to develop the "bi-component" system of genetically segregating Cas9 and the sgRNA *in vivo*, assessing the mutations at the target locus in pre-implantation, non-viable embryos.
4. In results chapter 4, I aim to generate stable transgenic embryonic stem cell and mouse lines, where a Cas9 transgene is integrated onto the X and Y sex chromosomes.

Chapter 2. Materials & Methods

2.1 Mouse Lines

2.1.1 Breeding and maintenance

All animals were maintained with appropriate care according to the United Kingdom Animal Scientific Procedures Act 1986 and the ethics guidelines of the Francis Crick Institute. All mice used were *Mus Musculus*. Wildtype mice were C57BL/6J and transgenic mice were maintained on a C57BL/6J background. When targeted embryonic stem cells were used to generate the sex chromosome-linked transgenic lines, the targeted lines were C57BL/6N and were injected into C57BL/6J albino embryos. Chimeric mice and germline transmitting genetically modified mice were then maintained on a C57BL/6J background. The TARGATT attPx3 mice were generated using C57BL/6N embryos and were backcrossed to C57BL/6J for at least seven generations prior to purchase from Charles River for zygotic microinjection at the Crick institute. The mice were kept in IVC cages with automatic watering systems and air management systems which controls air flow, temperature and humidity. Animals were checked on a daily basis by the Biological Research Facility (BRF) staff. All of the breeding units are Specific Pathogen Free (SPF).

2.1.2 Timed matings

Male and female mice were co-housed from approximately 17:00 by the addition of a male mouse to the female cage. The females were checked for the presence of a vaginal plug the following morning, each day, indicating a mating had taken place. If a plug was visible, this day was considered embryonic day (E) 0.5.

2.1.3 Embryo collections

Following the presence of a vaginal plug (see 2.1.2) plugged females were culled according to Schedule 1 Killing (S1K) methods, according to the required embryonic age. The uterine horns were dissected into dPBS (Dulbecco's Phosphate Buffered Saline, Gibco, ThermoFisher Scientific, USA) and individual horns flushed with 1ml of FHM (Follicle Holding Medium, Appendix A) using a 1ml syringe and needle, under a Leica

light dissecting microscope (Leica, Germany), into FHM in a clean dish. Embryos were transferred into individual drops of FHM using a Stripper® (Origio, Denmark) pipette.

2.2 Molecular Biology

2.2.1 Genomic DNA extraction

Genomic DNA was extracted using different methods depending on tissue or cell type and downstream processes.

2.2.1.1 Mouse ear biopsies

Genomic DNA was extracted from mouse ear biopsies for PCR genotyping. 150µl 50mM NaOH was added to biopsies and heated at 95°C for 90 minutes. The samples were cooled to room temperature before addition of 15µl 25mM Tris HCl pH8 and centrifugation for 5 minutes at 13,000rpm.

2.2.1.2 Embryonic stem cells from 96-well plates

Genomic DNA was extracted from embryonic stem cells (ESCs) for PCR genotyping of CRISPR-Cas9 targeted lines. The ESCs were grown to confluency in 96 well plates. Excess medium was removed by plate inversion, and the ESCs were washed once with 150µl dPBS. The dPBS was removed by plate inversion. The ESCs were lysed in 50µl Bradley Lysis buffer (Appendix A). Proteinase K (final concentration 1mg/ml) was added to the lysis buffer immediately before addition of the buffer to the ESCs. The plate was sealed with parafilm before heating in a humidified incubator overnight at 60°C. The following day, the plate was cooled to room temperature before DNA precipitation by addition of 100µl ice-cold EtOH/NaCl (100% EtOH 394ml, 5M NaCl 6ml). The plate was incubated at room temperature for 30 minutes, followed by centrifugation at 3,000rpm for 20 minutes. The plates were inverted to remove excess liquid and DNA pellets were washed with 150µl ice-cold 70% EtOH twice. The supernatant was removed by inversion. Plates were air-dried and the DNA resuspended in 30µl EB (elution buffer; Appendix A).

2.2.1.3 Embryonic stem cells or mouse tissue

To produce larger volumes of genomic DNA, phenol-chloroform DNA extraction was carried out. In general, phenol-chloroform extraction was used for >6cm plates of confluent ESCs, or approximately 100mg of mouse tissue (fresh, or snap frozen immediately upon collection). The ESCs were dissociated into single cells and centrifuged (as in 2.6.3) before resuspension in 705µl Bradley Lysis buffer (see 2.2.1.2) with Proteinase K. Tissue pieces were thawed on ice, macerated and resuspended in 705µl Bradley Lysis buffer with Proteinase K. The samples were incubated overnight at 55°C. The next day, 10µl of 100ug/ml RNase A was added to the samples for 1 hour and incubated at 37°C. Following RNase A treatment, 750µl phenol was added to the samples before rotation for 15 minutes. The samples were then centrifuged at 15,000g for 5 minutes. The upper phase was collected into a fresh tube and 750µl phenol-chloroform added, the samples inverted and centrifuged at 15,000g for 5 minutes. The upper phase was collected into a fresh tube and 750µl chloroform-isoamyl alcohol added, the samples inverted and centrifuged at 15,000g for 5 minutes. The upper phase was collected into a fresh tube, and 750µl isopropanol added, the samples inverted and centrifuged at 1,000g for 15 minutes. The DNA pellet was washed with 1ml of 70% EtOH twice, and dried at room temperature for approximately 5-10 minutes. The DNA was resuspended in EB.

2.2.2 Polymerase Chain Reaction (PCR)

Unless otherwise stated, PCR genotyping reactions were carried out in a total volume of 25µl using MyTaq Red Mix (2X, Bioline, UK) with standard thermocycling conditions according to manufacturer's instructions. Amplification of large products (>1kb), or if PCR amplicons were required for downstream cloning applications; Q5 Hot Start High-Fidelity DNA polymerase was used, according to manufacturer's instructions (New England Biolabs (NEB), USA).

Correct product amplification was confirmed by agarose (Agarose, Sigma-Aldrich, USA) gel electrophoresis. The agarose powder was melted in 1x TAE buffer (Tris-acetate-EDTA), cooled and stained with SybrSafe (ThermoFisher Scientific, USA) before casting. Purple loading dye (6X, NEB, USA) was added to PCR samples at a final

concentration of 1X for visualisation under ultraviolet light. The PCR products were run alongside a 1kb GeneRuler DNA ladder (ThermoFisher Scientific, USA). The agarose gels were run at approximately 90V for 45 minutes, or until clear band separation. PCR amplicons on the agarose gel were visualised using BioRad ChemiDoc XRS+ Imaging System. Primers (listed in Appendix B) were designed using publicly available tool Primer3 [http://primer3.ut.ee/;(Untergasser et al., 2012, Koressaar and Remm, 2007, Koressaar et al., 2018)] or previously published sequences, and used at a final concentration of 10µm. Unless otherwise stated, all primers were synthesised by Eurofins Genomics (Germany).

2.2.3 MiSeq PCR

Unless otherwise stated, MiSeq PCR reactions were carried out in a total reaction volume of 25µl using Q5 High-Fidelity 2X Master Mix (Qiagen, UK). MiSeq primers were designed using Primer3, with additional adaptor sequences (Appendix B), to amplify products <500bp and used at a final concentration of 5µm. PCR products were amplified under standard thermocyling conditions, according to manufacturer's instructions (Qiagen, UK). The MiSeq PCR amplicon was confirmed by gel electrophoresis before purification using solid phase reverse immobilisation (SPRI) beads (made in house). Downstream library amplification was carried out with Illumina TruSeq indexing primers (Illumina, USA; see 2.9.1 below).

2.2.4 RNA extraction

The ESCs were grown to confluency on pre-coated plates (see 2.6.1) in 2i+LIF culture conditions, as standard (see 2.6.3). Medium was aspirated, followed by washing with dPBS. The dPBS was aspirated and the adherent ESCs were lysed in 1ml TRI Reagent (Sigma-Aldrich, USA). The lysis was transferred to a microcentrifuge tube and centrifuged at 12,000g for 10 minutes at 4°C. The cleared upper phase was transferred to a new microcentrifuge tube and 200µl chloroform (Sigma-Aldrich, USA) added. The solution was mixed by shaking for 15 seconds, followed by centrifugation at 12,000g for 15 minutes at 4°C. The aqueous upper phase was transferred to a new microcentrifuge tube and RNA was precipitated by addition of 500µl isopropanol, incubation at room temperature for 10 minutes, followed by centrifugation at 12,000g for 10 minutes at 4°C.

The supernatant was removed and pellet washed with 75% EtOH, followed by centrifugation at 7,500g for 5 minutes at 4°C. The pellet was air dried before resuspension in RNase-free water.

2.2.5 cDNA synthesis

The RNA and reverse transcriptase reagents were thawed, briefly centrifuged and kept on ice until use. cDNA synthesis was performed according to cDNA synthesis manufacturer's instructions (Thermo Scientific Maxima First Strand cDNA Synthesis Kit, USA). For downstream qPCR reactions, 200ng of RNA was used per cDNA synthesis reaction. For downstream RT-PCR reactions, a reverse transcriptase minus (-RT) negative control was always included to confirm absence of genomic DNA contamination.

2.2.6 Quantitative PCR (TaqMan probes)

After cDNA synthesis, the 20µl product was diluted 1:10 in water. Unless otherwise stated, each qPCR reaction was made to a total volume of 10µl (5µl TaqMan 2X Universal PCR Master Mix (Applied Biosystems by ThermoFisher Scientific, USA), 0.5µl TaqMan probe (ThermoFisher Scientific, USA), 2.5µl water, 2µl of diluted cDNA). Each DNA sample was tested in triplicate and normalised to a housekeeping gene control, usually *Gapdh,* unless otherwise stated. All TaqMan probes used are listed in Appendix C.

2.2.7 TOPO XL cloning

Standard PCR reactions were done according to protocols for large PCR products, using Q5 Hot-Start High-Fidelity DNA polymerase (see 2.2.2). Presence of amplicons was initially confirmed by agarose gel electrophoresis. PCR products were used immediately for downstream TA cloning reactions according to manufacturer's instructions (TOPO® XL PCR Cloning Kit, ThermoFisher Scientific, USA).

2.2.8 Transformations and plasmid isolation

Chemically competent *E. coli* cells (90µl aliquot; made in house), were thawed on ice immediately prior to transformation. The cloning reaction was added directly to the

competent cells (made in house, UK) and incubated on ice for 30 minutes. Samples were then heated at 42°C on a heat block for 1 minute, followed by cooling on ice for 2 minutes. 810µl S.O.C (super optimal broth with catabolite repression, made in house; Appendix A) was added, and followed by heating in a shaking incubator for 30 minutes at 37°C, 200rpm. Pre-prepared LB agar (lysogeny broth) plates containing an antibiotic (either ampicillin, 100mg/ml or kanamycin, 50mg/ml) were pre-warmed at 37°C during this incubation step. Following shaking, the tubes were centrifuged at 1,000g for 5 minutes and the supernatant removed. The cell pellet was resuspended in minimal volume, spread on the plates and incubated overnight at 37°C. Individual colonies were picked the following day into 2ml of liquid LB with antibiotic (either ampicillin, 100mg/ml or kanamycin, 50mg/ml) and placed in a shaking incubator overnight at 37°C. The following day, the plasmid DNA was isolated using a mini-prep kit according to manufacturer's instructions (Qiagenprep Spin Miniprep Kit, Qiagen, UK). For larger quantities of plasmid, colonies were picked into 50-200ml of liquid LB with antibiotic and placed in a shaking incubator at 37°C overnight. The plasmid DNA was then isolated using a maxi-prep kit according to manufacturer's instructions (GenElute HP Plasmid Maxiprep Kit, Sigma-Aldrich, now Merck, USA).

2.3 Protein Biology

2.3.1 Protein extraction

Protein extraction buffer (in RIPA (radioimmunoprecipitation assay) buffer; both Appendix A) was pre-prepared and kept on ice until use. For protein extraction from ESCs, the cells were firstly lysed into a single cell suspension, before centrifugation and resuspension in 50-100µl protein extraction buffer. For protein extraction from mouse tissue, the tissue was collected on ice, cut into small pieces and snap frozen in liquid nitrogen in screw-cap tubes (thereafter kept at -80°C). A single piece of tissue (approximately 100mg) was thawed on ice, and resuspended in 100-200µl protein extraction buffer. For both ESCs and mouse tissue, the resuspended samples were kept on ice for 30-40 minutes with occasional mixing to ensure complete homogenisation, followed by centrifugation at 8,000rpm at 4°C for 10 minutes. The supernatant was transferred to a clean 1.5ml Eppendorf tube and kept on ice until use.

2.3.2 Protein quantification

Protein concentration was quantified using a bicinchoninic acid (BCA) assay using bovine serum albumin (BSA, 2mg/ml) serially diluted and the protein samples diluted 1:10 in duplicate. An Ensight Multimode Plate Reader, by PerkinElmer with software Kaleido 2.0 was used for colorimetric quantification.

2.3.3 Western blot

The protein samples (10-100µg) were diluted in water and Laemmli buffer (Appendix A) and heated to 98°C for 10 minutes, followed by brief centrifugation. The samples and ladder were loaded into gel wells, inserted into the BioRad tank and covered by running buffer (Appendix A). The gel was run at 50V until the samples left the wells, followed by running at 120-150V for 45-60 minutes. The membranes were prepared by soaking in transfer buffer (Appendix A). The gel and membranes were then inserted into the transfer tank and run at 60-140V for 45-90 minutes. The membranes were then blocked in blocking buffer (Appendix A) for 60 minutes. Following blocking, the membranes were incubated with a primary antibody (diluted in blocking buffer) overnight at 4°C. The following day, the membranes were washed 3x in TBS+0.1% Tween followed by incubation with a secondary antibody (diluted in TBS+0.1% Tween), at room temperature for 45-60 minutes. The membranes were washed 3x in TBS+0.1% Tween before soaking in ECL for imaging. All antibodies used in western blots are listed in Appendix D. Western blots were carried out by Valdone Maciulyte (senior LRS, Turner lab).

2.4 Southern Blot

Genomic DNA was extracted from either ESCs or mouse tissue by phenol-chloroform DNA extraction (see 2.2.1.3). Resulting genomic DNA (10µg) was digested overnight with restriction enzymes (listed in Appendix E) and purified the following day by standard phenol-chloroform DNA precipitation, and washing with 70% EtOH. Digested and precipitated DNA was loaded (10µg/lane with 6X loading dye at final concentration 1X) onto a 1% agarose gel (stained with SybrSafe). Bromophenol blue was loaded in one lane for monitoring later colour changes. The agarose gel electrophoresis was run at either 30V overnight or 50V for 5 hours.

The agarose gel was then depurinated by washing in depurination buffer (Appendix A) with gentle agitation until the bromophenol blue turned yellow (approximately 10 minutes). The agarose gel was rinsed in water, followed by washing in denaturation buffer (Appendix A) with gentle agitation for 30 minutes (the bromophenol blue turned blue). The denaturation buffer was then replaced with neutralisation buffer (Appendix A) and was washed with gentle agitation for 30minutes. The agarose gel was then rinsed with water to remove all traces of the neutralisation buffer.

A glass sheet was placed across a metal tray that contained 20X SSC. On top of the glass sheet, Whatman filter paper (3MM, Sigma-Aldrich) was placed, lying across the glass sheet and into the container acting as a wick. The neutralised agarose gel was then blotted by placing the gel upside down on top of fresh Whatman filter paper , on top of the glass sheet. On top of the gel and filter paper, was a positively charged nylon membrane, 3 filter papers and 2 inches of paper towels. A heavy weight (~1.5-2kg) was placed on top and the blot left to transfer overnight.

The following day, the DNA was fixed by cross-linking by placing 4 Whatman sheets in a tray with 10X SSC (enough to cover the sheets), followed by the membrane (DNA facing up) and the tray was placed in a Stratalinker UV crosslinker (auto cross-linking, 1200U joules, ~2 minutes). The membrane was then dried, rinsed twice in purified water and then dried again.

The probe was synthesised according to manufacturer's instructions (PCR DIG Probe Synthesis Kit, Roche), and 2µl run on 2% agarose gel to confirm DIG labelling, followed by purification using a PCR purification kit (e.g. Qiagen, UK), eluting in 30µl. For hybridisation, hybridisation buffer (Appendix A) was pre-heated to an optimal temperature (T_{opt}; according to DIG Easy Hyb manufacturer's instructions, Roche) for hybridisation of the specific probe. The membrane, with DNA facing the inside, was then placed inside a bottle containing pre-heated hybridisation buffer and incubated for 30 minutes at T_{opt}. The probe was diluted in 50µl purified water and boiled at 98°C for 10 minutes. The probe was then cooled quickly on ice for 10 minutes. Salt was added to the

ice to cool the probe faster. The pre-heated hybridisation buffer was added to the probe (0.5-4µl) in a separate container. The pre-hybridisation solution in the bottle containing the membrane was discarded and quickly replaced with the hybridisation buffer containing the probe. The bottle was then incubated in the hybridisation oven at T_{opt} overnight.

The following day, the membrane was rinsed twice in 2X SSC/0.1% SDS for 5 minutes at room temperature. The membrane was then washed twice in 0.1X SSC/0.1% SDS at 65°C for 15 minutes, in the hybridisation bottle, using the hybridisation oven. The membrane was then rinsed in water.

For blocking and detection, 10X blocking buffer (Appendix A) was filtered with a 0.45µm filter disc and used to make 1X blocking buffer in maleic acid. The membrane was incubated in 1X blocking buffer, with gentle agitation for 30 minutes (up to 3 hours). The membrane was then incubated in antibody solution (DIG antibody in 1X blocking buffer, 1:50,000). The membrane was then washed twice in washing buffer (Appendix A) for 15 minutes. The membrane was then equilibrated in detection buffer, for 5 minutes on a shaker. CSPD (a chemiluminescent substrate, Roche) was diluted 1:1000 in 2ml detection buffer, and applied to the membrane for 5 minutes. The film was developed in a dark room to visualise the DNA bands. All southern blots were carried out by Valdone Maciulyte.

2.5 Preparation of CRISPR Components

After generation of CRISPR components, the resulting plasmids were transformed into chemically competent *E. coli* cells, the plasmid isolated (see 2.2.8) and Sanger sequenced (Genewiz, UK).

2.5.1 sgRNA design, cloning and synthesis

The 20mer crRNA component of each sgRNA sequence was designed using publicly available tools (crispr.mit.edu, Zhang lab) and sgRNAs with a predicted high on-target activity were selected. Oligonucleotides, with the addition of BbsI or BsaI overhangs, (synthesised by Eurofins Genomics, Germany) were annealed and ligated into sgRNA-

mCherry ("pLethal"), pX330, a gift from Feng Zhang, Addgene plasmid #42230 (Cong et al., 2013), pX459v2 a gift from Feng Zhang, Addgene #62988, (Ran et al., 2013b) and/or pX458, a gift from Feng Zhang, Addgene #48138, (Ran et al., 2013b) using BbsI or BsaI. All oligonucleotides are listed in Appendix B.

2.5.2 Targeting vectors

2.5.2.1 Cas9-eGFP

To generate the *Hprt* X-Cas9 (X chromosome) and Y-Cas9 (Y chromosome) targeting vectors whereby expression of Cas9 and eGFP is linked via a T2A peptide under a CMV early enhancer/chicken β-actin (CAG) promoter, pX458 (Ran et al., 2013b) was modified. The 5' and 3' homology arms were cloned using unique restriction enzyme sites by standard directional cloning techniques. The LoxP-PGK-neomycin-LoxP cassette was inserted using Gibson assembly using unique restriction sites according to manufacturer's instructions (NEBuilder HiFi DNA Assembly Cloning Kit, NEB, USA).

2.5.2.2 sgRNA-mCherry ("pLethal")

To generate the sgRNA and mCherry expressing plasmid, pX333, a gift from Andrea Ventura, Addgene #64073 (Maddalo et al., 2014), containing two tandem U6 promoters was modified. The Cas9 gene sequence was replaced by the mCherry sequence using standard directional cloning techniques (AgeI and EcoRI restriction enzymes), under a Cbh promoter. The sgRNA oligonucleotides (Appendix B) were cloned into the pLethal plasmid using restriction enzyme sites BbsI and BsaI for U6 promoter 1 and 2, respectively.

2.5.2.3 attB-sgRNA-mCherry-attB (TARGATT)

The tandem U6-sgRNA cassettes and Cbh-mCherry components of the pLethal plasmid were cloned into TARGATT plasmid #3 (Applied StemCell, USA; (Zhu et al., 2014, Tasic et al., 2011) using restriction enzyme sites SpeI, XbaI and NotI by standard directional cloning. The sgRNA-mCherry sequence was then flanked by attB sequences contained in the TARGATT plasmid for integrase mediated recombination into the H11-attPx3 mouse line (Applied StemCell, USA; (Zhu et al., 2014, Tasic et al., 2011).

2.6 Embryonic Stem Cell Culture

2.6.1 Preparation of cell culture plates

All ESCs were plated on laminin unless otherwise stated. Tissue culture grade plasticware (Nunc, ThermoFisher Scientific, USA) was coated with poly-L-ornithine (0.01% in water) for 1 hour (2.5ml per 6cm plate) at 37°C. The solution was aspirated, plates washed twice with 1x dPBS (1ml per 6cm plate), followed by addition of laminin solution (10ng/ml in dPBS; 2.5ml per 6cm plate) for 1 hour at 37°C (Hayashi and Saitou, 2013). The solution was aspirated immediately prior to plating ESCs.

2.6.2 Deriving embryonic stem cell lines

Blastocyst-stage embryos were collected (see 2.1.3 above) and plated individually in wells of a 24-well plate in 500µl pre-warmed 2i+LIF (Appendix A) and cultured in a 37°C, 5% CO_2 incubator. After 1 week, a further 1ml 2i+LIF was added. After approximately 3 weeks, any expanded blastocysts were dissociated into single cells by placing in 15µl TrypLE (Gibco, ThermoFisher Scientific, USA) and incubation for 4 minutes at 37°C , followed by quenching with 2i+LIF and dissociation by gentle pipetting. The single cell suspension was plated into one well of a pre-coated (see 2.6.1 above) 4-well plate and passaged as in 2.6.3.

2.6.3 Passaging and maintenance

The required number of plates were coated ahead of time (see 2.6.1 above) and the 2i+LIF pre-warmed. The 2i+LIF from plates containing actively growing ESCs was aspirated. The adherent ESCs were washed once with 1x dPBS, followed by addition of TrypLE (1ml per 6cm plate) and incubated at 37°C for 3 minutes, until all the ESCs had detached. The TrypLE was quenched by addition of 2ml 2i+LIF (1:2 TrypLE:2i+LIF), and the ESCs were dissociated into single cells by gentle pipetting. The resulting ESC suspension was cell-counted using the EVE hemocytomer (EVE by Cambridge Bioscience, UK). The required volume of cell suspension was centrifuged at 200g for 3 minutes, the supernatant aspirated, and the pellet resuspended in 1ml 2i+LIF before plating with further 2i+LIF and incubating at 37°C, 5% CO_2. The ESCs were passaged every 2-3 days to prevent

occurrence of abnormal karyotypes and generally passaged at a ratio of 1:4 plates at each new passage.

2.6.4 Freezing

Freezing solution (Appendix A) was made and chilled at 4°C prior to freezing ESCs. The ESCs were dissociated into single cells and centrifuged (as in 2.6.3 above. The supernatant was aspirated and ESCs resuspended in 250µl per cryovial. The freezing solution was added to the resuspended cells at a ratio of 1:1 (250µl cell suspension: 250µl freeze solution). The total volume (500µl) was transferred into each cryovial and kept at -80°C in freezer boxes overnight. The following day, the cryovials were transferred to liquid nitrogen for long term storage.

2.6.5 Thawing

Plates were coated ahead of time (see 2.6.1) and the 2i+LIF pre-warmed. The cryovials containing frozen ESCs were warmed at 37°C for approximately 2-3 minutes, until visibly thawed. Once thawed, 1ml 2i+LIF was added to each cryovial, drop-wise. The cell suspension was transferred to a 15ml falcon tube, and a further 4ml 2i+LIF added slowly. The cell suspension was centrifuged at 200g for 4 minutes, the supernatant aspirated, the pellet resuspended in 1ml 2i+LIF and ESCs plated.

2.6.6 Fluorescence Activated Cell Sorting (FACS)

The ESCs were dissociated into a single cell suspension and centrifuged (see 2.6.3 above). The ESC pellet was then resuspended in FACS medium (Appendix A) and the cell suspension passed through a 40µm filter. The ESCs were sorted on the Aria Fusion Cell Sorter.

2.6.7 Standard transfections

In general, $1x10^6$ single cells were plated on pre-coated 6-well plates in 2i+LIF (see 2.6.1 and 2.6.3). Following plating, C57BL/6J or C57BL/6N background ESCs were transfected with Lipofectamine 3000 (ThermoFisher Scientific, USA) according to manufacturer's instructions (Opti-MEM by ThermoFisher Scientific, USA), with 1µg of

plasmid (typically at concentration 1µg/µl) per well. The lipofection reaction was incubated for 15 minutes at room temperature before addition to plated cells. The CRISPR-Cas9 transfections were based on published protocols (Ran et al., 2013b).

2.6.8 Generating knock-in lines by CRISPR-Cas9 homology directed repair (HDR)

For each targeting experiment, $2x10^6$ C57BL/6N (line B6N6; generated in-house by GeMS) single ESCs were transfected using Lipofectamine 2000 (ThermoFisher Scientific, USA; 22µl lipofectamine 2000, 478µl Opti-MEM, according to manufacturer's instructions) with the addition of 2ug pX330 sgRNA expressing plasmid, and 2ug of repair template plasmid (targeting vector). The solution was incubated at room temperature for 20 minutes prior to addition to cells. The lipofectamine-DNA mix was added to C57BL/6N ESCs, prepared as a single cell suspension, and incubated at room temperature for 10 minutes before seeding ESCs in 6-well plates pre-coated with feeders on gelatin. To select for transgenic clones, 2 days post-transfection, neomycin antibiotic was added (270mg/ml) for 8-10 days. The X-Cas9 (X chromosome targeting) and Y-Cas9 (Y chromosome targeting) knock-in transfections on serum+LIF conditions were done by the Genetic Manipulation Service (GeMS), part of the Science Technology Platforms at the Francis Crick Institute.

2.7 Microinjections

2.7.1 Zygote (TARGATT)

H11-attPx3 TARGATT female mice (Applied StemCell, USA, (Zhu et al., 2014, Tasic et al., 2011) were superovulated and oocytes harvested. The oocytes were *in vitro* fertilised by sperm collected from C57BL/6J male mice. The TARGATT targeting vector (attB-sgRNA-mCherry-attB) and C31ϕ integrase were microinjected into the pronuclei of zygote stage embryos. The embryos that successfully divided to the 2-cell stage were transferred into pseudopregnant females. All pronuclear microinjection and surgical work was undertaken by the GeMS.

2.7.2 Blastocyst

Correctly targeted ESCs (see 2.6.8) were trypsinised into a single cell suspension and resuspended in 100µl KSOM (Potassium supplemented simplex optimised media, Appendix A; Lawitts and Biggers, 1993). Approximately 10 single cells were injected into blastocyst-stage albino C57BL/6J embryos. The injected blastocysts were then transferred to pseudopregnant females. Chimerism of resulting pups born was assessed by percentage black coat colour. All blastocyst microinjection and surgical work was undertaken by the GeMS.

2.8 *In Vivo* Imaging

Mouse pups were imaged for presence of fluorescence transgenes at age 3-4 days using the IVIS Lumina XR (Caliper LifeSciences) with "Living Image 4.4" software. To investigate expression of mCherry positive pups, the excitation filter was set to 535nm and the emission filter to dsRed.

2.9 Next Generation Sequencing (NGS)

2.9.1 MiSeq library preparation

The purified MiSeq amplicon (see 2.2.3 above) was carried forward for library preparation, according to the Illumina MiSeq library prep manufacturer's instructions (Nextera Index Kit V2). The resulting indexed library then underwent a second round of purification using Agencourt AMPure beads (Beckman Coulter, USA). The purified library was then quantified, normalised and pooled before submission to the Advanced Sequence Facility (ASF, part of the Science Technology Platforms at the Crick institute) for sequencing on the MiSeq for 2x250bp sequencing. The library indexing, purification and pooling steps were carried out by the GeMS.

2.9.2 MiSeq data analysis

The resulting Fastq files were collapsed using the FastX Toolkit (v0.0.13) and aligned to the reference genome (*Mm10*) using blastn, by the ASF and Bioinformatics and Biostatistics facility (BABS, part of the Science Technology Platforms at the Crick

institute). The MiSeq reads were analysed using a previously published R package; CrispRVariants (Lindsay et al., 2016) to evaluate individual sgRNA mutation efficiencies. The bioinformatic analysis of the MiSeq reads in order to generate the mutation efficiency was carried out by Jasmin Zohren (post-doc, Turner lab).

2.9.3 Whole genome sequencing (WGS) library preparation

Genomic DNA was extracted from embryonic stem cells by phenol-chloroform DNA extraction (see 2.2.1.3). A quality control (QC) step was performed using Tapestation Analysis Software 3.1 (Agilent Technologies) by the ASF. Samples that passed the QC step were then carried forward for library preparation, using the KAPA library preparation kit (KAPA Biosystems, Roche) according to manufacturer's instructions. The library preparation step was carried out by the ASF.

2.9.4 Whole genome sequencing data analysis

The whole genome sequencing FastQ reads were trimmed using Trim Galore! to remove adaptor sequences. Post-trimming, the reads were aligned using BWA. To analyse the karyotype of targeted embryonic stem cells, a previously published R package called QDNAseq (Scheinin et al., 2014), version 1.8.0 was utilised. The bioinformatic analysis of the low-pass whole genome sequencing data for karyotyping was performed by Jasmin Zohren.

2.9.5 RNAseq library preparation

Blastocyst-stage embryos were collected (in accordance with 2.1.3 above) and washed 3 times in FHM droplets, and once in dPBS and placed in individual low-bind RNase-free 0.2ml PCR tubes in minimal volume. The cDNA synthesis, amplification and purification steps were performed according to manufacturer's instructions, using the SMART-Seq® v4 Ultra® Low Input RNA Kit (Takara Bio USA Inc, USA). A QC step was performed by the ASF using the Agilent 2100 Bioanalyzer. Samples that passed QC were carried forwards for library preparation according to manufacturer's instructions (Nextera XT DNA Library Preparation Kits, Illumina, USA). The Illumina library preparation was carried out by the ASF. The libraries were sequenced on the HiSeq 4000. The minimum number of reads returned per embryo was 12.5 million, 100bp paired-end reads.

2.9.6 RNAseq data analysis

The quality of the sequencing for all samples was assessed by MultiQC. Failed samples were removed from the analysis. A total of 41 samples were carried forward for further analysis. The RNA sequencing FastQ reads were mapped using HISAT2 (Kim et al., 2015). The default HISAT2 parameters were used, which allowed for soft-clipping of reads without trimming. The .sam files were converted into .bam files and indexed using SAMtools (Li et al., 2009). The RNAseq reads were counted and annotated using the RSubRead function "featureCounts" (Liao et al., 2014). For each sample I generated matrices containing each sample and gene feature information, including genes, exon-exon boundaries, and promoters. I used the inbuilt annotation matrix for the mouse genome *Mm10*. I changed the default featureCounts parameters to account for paired-end reads:

```
MyfeatureCounts <- featureCounts(vectorfile, annot.inbuilt =
"mm10", isPairedEnd = TRUE)
```

The annotation and count matrices generated by featureCounts were carried forward for differential gene expression (DESeq) analysis. Statistical analysis of differential gene expression was performed in RStudio Version 1.2.1335 using DESeq2 (Love et al., 2014). I performed DESeq analysis for individual genes of interest. I generated plots of individual differential gene expression using ggplot2. I adapted the default parameters for aesthetics, and to show the mean expression value (example shown for *Uty* expression):

```
uty <- plotCounts(dds, gene = "22290", intgroup = "condition",
returnData = TRUE)

sp_uty <- ggplot(uty, aes(x = condition_sex, y = count, color =
condition)) + geom_point(position=position_jitter(w = 0.1,h =
0), size=6) + ggtitle("Uty") + theme(axis.title.x =
element_text(size=12),axis.title.y = element_text(size=12))

sp_uty + stat_summary(fun.y=mean, aes(group="wt"), geom="point",
colour="black", size=15, shape=95)
```

I generated the PCA and heatmap with `rlog` transformed log2 counts, according to the DESeq2 vignette.

Chapter 3. Results 1: Female specific lethality by expression of Cre inducible Diphtheria Toxin A

3.1 Introduction

Generating single-sex litters could have a positive impact on many areas of biology, including health, agriculture and in biomedical or scientific research, as described in section 1.6. When harnessing a method that relies on embryonic non-viability of the non-required sex, there must be no negative effect on the development of the required sex.

In this Chapter I generated a proof-of-principle experiment to determine whether sex-specific embryonic non-viability could be genetically induced. Further, I examined whether the survival of the non-affected sex was compromised. I firstly used multiple pre-existing mouse models to investigate the possibility of generating single-sex litters. Secondly, I investigated the transcriptomic differences between surviving embryos from a mating where the non-required sex was embryonic lethal, compared to wildtype embryos.

In order to generate embryonic non-viability, the mouse model Gt(*ROSA*)26Sor$^{tm1(DTA)Kio}$ (Marques et al., 2009), hereafter referred to as "*R26*-DTA" was utilised. The *R26*-DTA transgene encodes an attenuated toxin; Diphtheria Toxin A (DTA) inserted into the permissive locus *Rosa26* (Friedrich and Soriano, 1991). DTA expression is inhibited due to a LoxP-STOP-LoxP cassette preceding the DTA sequence (Sternberg and Hamilton, 1981). Cre recombinase-excision of the floxed-STOP cassette induces DTA expression. Cre-induced DTA allows for precise spatiotemporal DTA expression. Previous studies have shown that expression of DTA in different tissue types results in cell death. For example, Marques *et al* crossed the *R26*-DTA mice with a *Tnfrs4* promoter driving Cre recombinase to cause T-cell specific ablation (Marques et al., 2009).

To induce DTA expression I wished to use a Cre recombinase that was expressed during early mouse gestation to maximise the length of time that DTA would be expressed.

Furthermore, for the Cre-induced DTA to be sex-specific, the Cre allele needed to be sex chromosome-linked. Therefore, I utilised the pre-existing mouse model *Hprt*$^{tm1(CAG-Cre)Mnn}$, hereafter "X-Cre" (Tang et al., 2002), where a constitutively active Cre cassette was inserted into the X-linked *Hprt* locus. In humans, *Hprt* has an essential function, with abnormal reduction in protein levels resulting in error of purine metabolism and excessive uric acid. The clinical diagnosis of this disease is called Lesch-Nyhan disease (Lesch and Nyhan, 1964). The symptoms include severe neurological disorders, including cognitive and attention deficits (Torres and Puig, 2007). Hemizygous men are more commonly affected whereas women are usually heterozygous and carriers for the disease. Conversely, *Hprt* loss-of-function in the mouse has no known detrimental phenotypes in viability, fertility and behaviour (Kuehn et al., 1987, Hooper et al., 1987, Koller et al., 1989, Jinnah et al., 1990). The lack of detrimental phenotype upon *Hprt* knock-out is seen in other species, such as male rats (Meek et al., 2016).

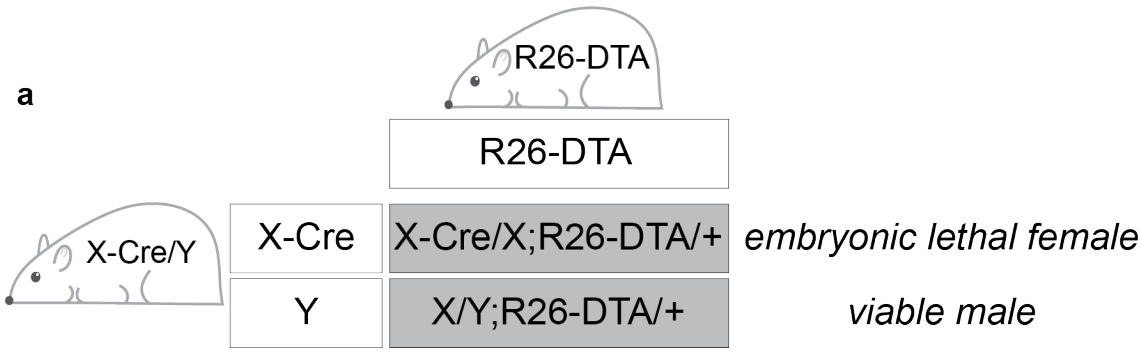

Figure 6. Hemizygous X-Cre/Y male mating to a homozygous *R26*-DTA female

(a) Female-specific co-inheritance of the X-linked Cre allele and *R26*-DTA allele results in Cre-induced activation of the DTA. Activation of the DTA results in female-specific embryonic non-viability. The males, which do not inherit the Cre allele, are unaffected.

If the father is hemizygous X-Cre, the transgene will be uniquely inherited by daughters. Therefore, when an X-Cre father is crossed with a homozygous *R26*-DTA mother, only the daughters will co-inherit the transgenes (Figure 6). Furthermore, co-inheritance of X-Cre and *R26*-DTA would induce DTA expression in a female-specific manner. I hypothesised that co-inheritance of X-Cre and *R26*-DTA would therefore result in female-specific lethality and all-male litters (Figure 6).

3.2 Results

3.2.1 Timing of Cre recombinase expression

Although *Hprt* is an extremely useful safe-harbour locus, there is a caveat to generating X-linked transgenic lines. This major caveat is dosage compensation of the X chromosomes by XUR and XCI. In the mouse pre-implantation embryo, XCI is imprinted, and the paternally-inherited X chromosome is silenced. The Xp XCI is retained in the extra-embryonic lineages such as the placenta. The epiblast, which contributes to the embryo-proper, undergoes X-reactivation before random XCI ensues, where either the maternal or paternal X is silenced (Takagi and Sasaki, 1975, Okamoto et al., 2004, Okamoto et al., 2005, West et al., 1977, Marahrens et al., 1997). The process and mechanism of imprinted XCI is discussed further in 1.3.4. An X-Cre transgenic father passes the X-Cre allele to the XX daughters. Therefore, the X-Cre transgene will likely be subject to imprinted Xp XCI. If silenced, the X-Cre transgene will remain transcriptionally inactive in the extra-embryonic lineages.

Therefore, I first aimed to determine the timing of expression of the Cre recombinase allele to indicate the expected timing of the floxed-STOP removal in the DTA mice. To answer this question, the pre-existing mouse line B6:129S6-Gt(*ROSA*)26Sor$^{tm14(CAG-TdTomato)Hze}$/J (hereafter "LSL-tdTom") was utilised. This mouse line is a Cre reporter that contains a LoxP-STOP-LoxP cassette upstream of a TdTomato fluorescent reporter (Madisen et al., 2010). The TdTomato reporter is expressed under a constitutive CAG promoter, however expression is inhibited by the presence of the floxed STOP cassette; thereby preventing TdTomato expression. Removal of the STOP cassette by expression of Cre recombinase induces TdTomato expression. In this strategy, TdTomato expression is controlled spatiotemporally by the Cre recombinase driver.

Timed matings were set up between X-Cre males and homozygous LSL-tdTom females (Figure 7a). I harvested embryos at each day of pre-implantation development (E1.5, E2.5 and E3.5; Figure 7b) to pinpoint the day of TdTomato activation by expression of Cre

recombinase. I hypothesised that approximately half of the offspring would be TdTomato positive. The TdTomato positive embryos would be female embryos, where inheritance of paternal X-encoded Cre removed the floxed-STOP cassette and induced TdTomato expression. Male embryos would be TdTomato negative. Fluorescence microscopy visualisation of the TdTomato reporter at E2.5 (8-cell stage) and E3.5 (blastocyst-stage) confirmed that Cre is expressed during pre-implantation development (Figure 7c). TdTomato expression was not detectable at E1.5 (2-cell stage; Figure 7c). Embryos collected at E1.5 and E2.5 were kept in KSOM culture at 37°C, 5% CO_2 until E3.5-4.0. At E3.5-4.0, embryos were confirmed again for TdTomato expression prior to PCR genotyping for the Y-linked gene, *Sly*. Amplification of *Sly* indicated whether the embryos were male (XY) or female (XX). The extended *ex vivo* development in KSOM until the E3.5-4.0 stage allowed me to detect TdTomato expression from embryos that could not be successfully phenotyped at E1.5. All TdTomato positive embryos (as determined by fluorescence microscopy phenotyping) were female (n=10/10, 100%), and all TdTomato negative embryos were male (n=5/5, 100%) aligning with expected results.

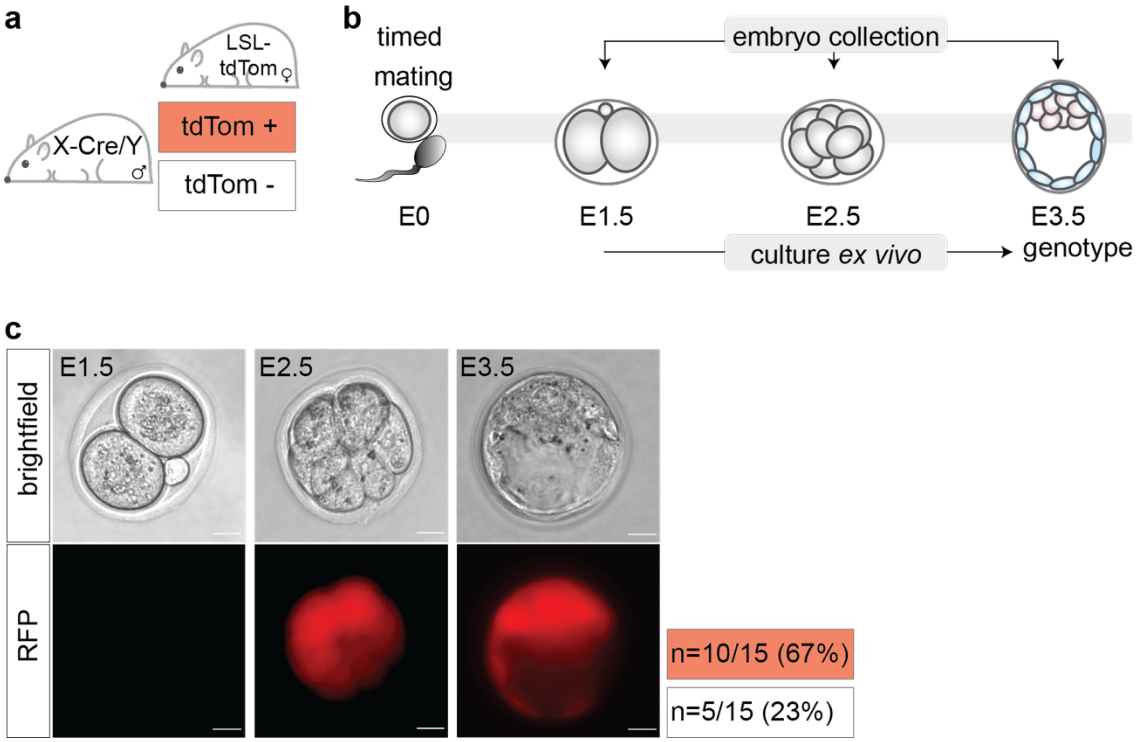

Figure 7. X-Cre induced TdTomato expression in female embryos

(a) Timed matings of X-Cre/Y males with homozygous LSL-tdTom females. (b) Pre-implantation stages of embryo development were collected at three time points; E1.5 (2-cell), E2.5 (~8-cell), and E3.5 (early blastocyst). (c) Pre-implantation embryos were phenotyped by fluorescence microscopy for expression of the tdTomato reporter. Scale bars=20µm.

Overall, this data confirmed that the Cre recombinase allele was inherited only by female offspring and expression of Cre-induced TdTomato was detectable by E2.5. Therefore, the X-Cre mouse line was appropriate to use for generating single-sex litters by Cre activation of a lethal toxin, DTA.

3.2.2 Cre recombinase-induced female-specific embryonic lethality

To determine if there was female-specific lethality resulting from Cre recombinase-induced DTA expression, experimental and control matings were set up and the pups born were genotyped at post-natal day 14 (PN14) for the Cre and DTA alleles. The control mating was X-Cre/Y males mated to wildtype (+/+) females. The experimental matings were either X-Cre/Y males mated to hemizygous *R26*-DTA/+ females, or homozygous *R26*-DTA/*R26*-DTA females. Pups were genotyped from multiple mating pairs, to ensure that the results were not due to mating-pair specific effects.

In the control matings, 10 litters were born, with a total of 56 pups (Figure 8a). Male and female pups were present (n=27, n=29 respectively), assessed by phenotypic examination. I expected only two possible genotypes; female X-Cre/X;+/+ and male X/Y;+/+ pups. The offspring were born at approximately the expected 1:1 sex ratio with females making up 52% of the total litters (Figure 8d). This data confirmed the Cre recombinase allele has no impact on the offspring sex ratio.

I next used X-Cre/Y males mated to heterozygous *R26*-DTA/+ females (Figure 8b). Mendelian inheritance predicted that there were four possible offspring genotypes: female X-Cre/X;+/+ (Figure 8b pale green), female X-Cre/X;*R26*-DTA/+ (Figure 8b dark green), male X/Y;+/+ (Figure 8b pale blue), male X/Y;*R26*-DTA/+ (Figure 8b dark blue). I hypothesised that co-inheritance of the X-Cre and *R26*-DTA allele would result in activation of the DTA allele, inducing embryonic non-viability. Therefore, I predicted that one-quarter of the litter would be lost prior to birth. In total, 13 litters were born from the heterozygous matings with a total of 52 pups. The pups were sexed by phenotypic analysis and ear notch biopsies were PCR-genotyped. At PN14, 65% (n=34/52) were male and 35% (18/52) were female. The ratio of male:female offspring was 2:1. As expected, two male genotypes were present; X/Y;+/+ (n=15, 44%) and *R26*-DTA/Y (n=19, 56%). Of the 18 female pups born, all were X-Cre/X;+/+ (18/18, 100%; Figure 8b). This showed that co-inheritance of the X-Cre and *R26*-DTA alleles induces embryonic lethality uniquely in females. The deviation from expected offspring genotype ratios was statistically-significant (Chi-squared test, $p=4x10^{-4}$).

Next, I investigated whether X-Cre males crossed with a homozygous *R26*-DTA female would induce a complete loss of female offspring, generating a single-sex litter (Figure 8c). First, pups were sexed by phenotypic examination. From nine litters born, with a total of 34 pups, all of the pups were male (34/34, 100%; Figure 8c). The complete sex ratio skew to all-male litters was a statistically-significant deviation from expected offspring sex ratios (Chi-squared test, $p=6x10^{-9}$; Figure 8d).

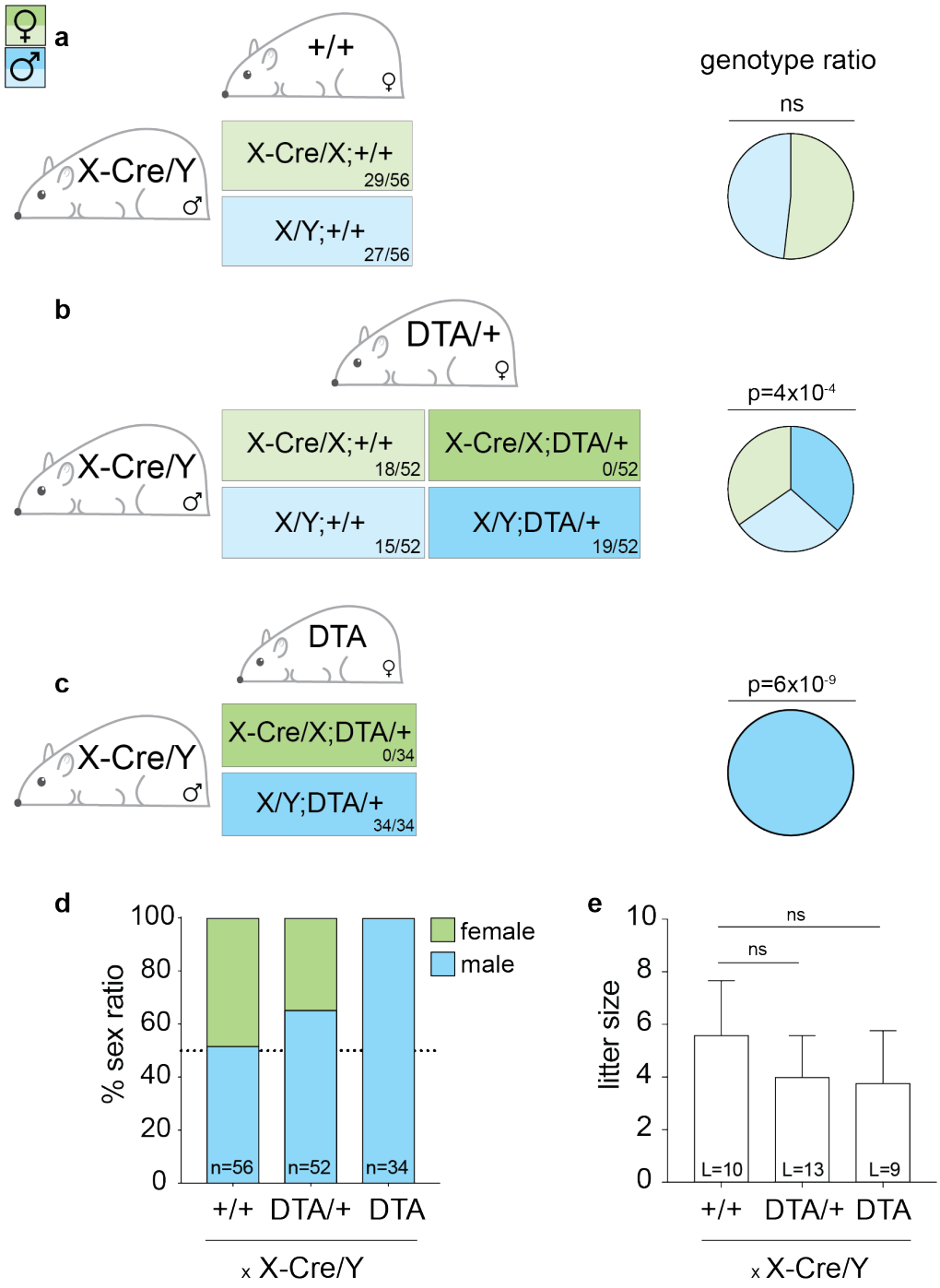

Figure 8. X-Cre induced Diphtheria toxin A female-specific lethality

(a) X-Cre/Y mating to C57BL/6J females (+/+). Deviation from expected offspring genotype ratios was not significant (p=1) (b) X-Cre/Y mating to heterozygous *R26*-DTA/+ females. Deviation from expected offspring genotype ratios was statistically-significant ($p=4x10^{-4}$). (c) X-Cre/Y matings with homozygous *R26*-DTA females. Deviation away from expected offspring genotype ratios was statistically-significant ($p=6x10^{-9}$). All statistical analysis performed using Chi-squared tests. (d) The female:male sex ratio (%) of offspring in litters for each mating. (e) Mean litter size for each mating. There was no statistically-significant reduction in mean litter size compared to the control X-Cre/Y mating. Statistical analysis performed using a Mann-Whitney test.

3.2.3 Effect of female lethality on litter sizes

Due to the loss of a quarter of the offspring in the heterozygous cross and half of the offspring in the homozygous cross, I wished to determine if there was an effect on litter sizes (Figure 8e). I hypothesised that in the heterozygous matings, the mean litter size would be reduced by approximately one-quarter compared to the control matings, whilst in the homozygous matings, I predicted the mean litter size should be reduced by half.

The mean litter size of the control mating was 5.6. According to my hypothesis, I predicted a mean litter size of 4.2 in the heterozygous cross. Results showed a mean litter size of 4.0 in this heterozygous mating, aligning with the expected reduction in litter size. The predicted mean litter size in the homozygous mating was 2.8. Interestingly however, the mean litter size from the homozygous cross was 3.8, i.e. greater than expected. Despite the reduction in litter sizes in the heterozygous and homozygous cross, there was no statistically-significant difference between the experimental litter sizes compared to the control.

3.2.4 Transcriptomic analysis of blastocysts

PCR genotyping of offspring born from the heterozygous (Figure 8b) and homozygous (Figure 8c) matings showed that female offspring that co-inherit the X-Cre and *R26*-DTA transgenes did not survive. The X-Cre allele was active by E2.5, as shown by TdTomato reporter expression, therefore I predicted that the *R26*-DTA allele was also active by this stage, and embryonic non-viability may occur around this time. To determine if the *in utero* embryonic lethality of the X-Cre/X;*R26*-DTA/+ females affected the surviving male littermates at the transcriptome level, I performed RNA sequencing (RNAseq) of blastocyst-stage embryos. I used the heterozygous mating because there were four possible offspring genotypes, one of which is wildtype, male X/Y;+/+. This genotype acts as an internal control, and allows for direct comparison of the transcriptome to wildtype blastocyst embryos from control matings. Experimental embryos were collected from X-

Cre/Y by *R26*-DTA/+ matings, while control embryos were from X-Cre/Y by C57BL/6J matings.

After blastocyst collections, the embryos were lysed, RNA-extracted, cDNA synthesis performed and libraries prepared. The resulting libraries were sequenced on the HiSeq 4000. The minimum number of paired-end reads per embryo was 12.5 million. The sequencing reads were mapped using HISAT2 (Kim et al., 2015) using default parameters, without trimming and allowing for soft-clipping of RNAseq reads. The resultant .sam files were converted into .bam files for downstream processes using SAMtools (Li et al., 2009). Next, the RNAseq reads were aligned to the mouse *Mm10* genome (Figure 9a). I used the Rsubread packaged with featureCounts function (Liao et al., 2014) to count the number of RNAseq reads and align to the target Mm10 genome to features such as genes, exons and promoters. RSubRead/featureCounts aligns paired-end, exon-containing and exon-exon reads, and performs soft-clipping for non-aligning nucleotides. To perform statistical analysis of differential gene expression, I used the featureCounts read count matrix and annotation matrix output for analysis using DESeq2 (Love et al., 2014). A detailed description of the RNAseq analysis protocols can be found in section 2.9.6.

To determine if it was possible to distinguish different embryo genotypes, I performed gene expression analysis for individual genes in DESeq2. The relative expression per gene was quantified by log2-fold change in normalised read counts for experimental samples versus control samples. Firstly, I genotyped blastocysts for expression of Y-linked genes *Uty* and *Eif2s3y*. Blastocysts that expressed the Y-linked genes were considered male (Figure 9b,c). To confirm the sex-genotyping, I analysed expression of *Xist*, which occurs only in females. *Xist* expression was detected in the non-Y-gene expressing blastocysts (Figure 9d). I next investigated three known blastocyst-expressed genes; *Pou5f1* (Chazaud and Yamanaka, 2016), *Cdx2* (Strumpf et al., 2005, Ralston and Rossant, 2008, Niwa et al., 2005) and *Gata6* (Chazaud et al., 2006), which are markers of the inner cell mass/epiblast, trophectoderm and primitive endoderm, respectively. All of the blastocysts expressed all three lineage specifiers (Figure 9e,f,g). Therefore, I concluded that the blastocysts analysed could be determined by the mating condition

(experimental versus control), genotyped by sex (male versus female), and expressed the expected embryonic lineage markers.

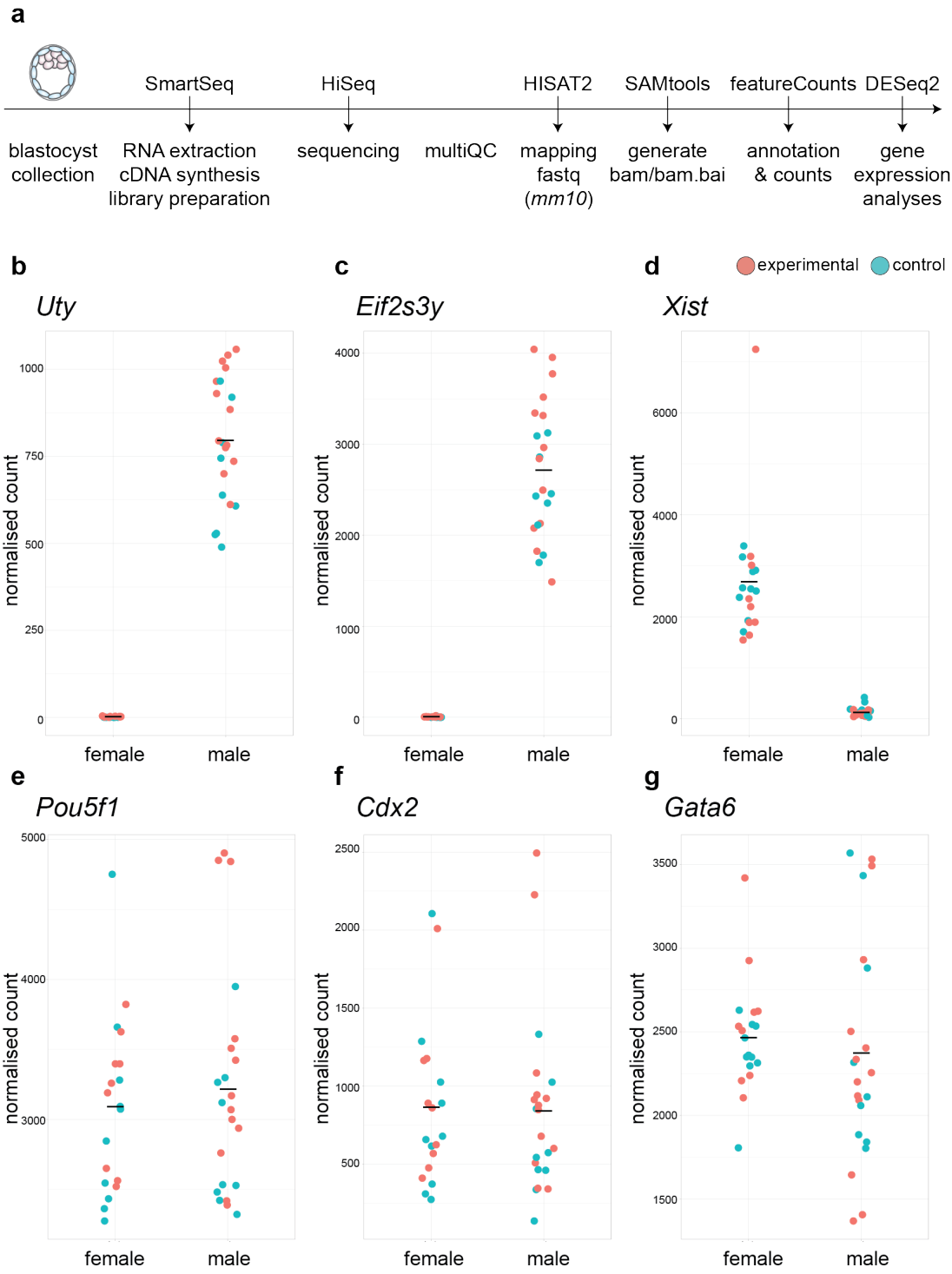

Figure 9. Relative gene expression

(a) Pipeline for generating gene expression analyses by DESeq2. (b-g) Relative gene expression between male and female blastocysts (b) *Uty* (c) *Eif2s3y* (d) *Xist* (e) *Pou5f1* (f) *Cdx2* (g) *Gata6.*

The library preparation and sequencing by HiSeq 4000 steps in the protocol were performed by the ASF STP.

Once I had confirmed that the blastocysts could be genotyped by sex, I examined expression of *Hprt* and *Rosa26,* the endogenous loci disrupted by transgene integration of X-Cre and *R26*-DTA, respectively. Integration of the X-Cre transgene into *Hprt* results in loss-of-function of *Hprt*. However the X-Cre is a paternal X-linked transgene, therefore Cre expression would be silenced in all female embryos, resulting in single maternal X-linked *Hprt* expression. Males carry one X chromosome and therefore only express *Hprt* from the single X. Therefore, I hypothesised that *Hprt* expression should be similar between different sexes and genotypes. Results of the relative gene expression analysis showed that *Hprt* expression was similar between experimental and control blastocysts, and within the experimental and control populations the relative expression did not differ between females and males (Figure 10a). These results aligned with what I expected for *Hprt* expression.

Integration of the *R26*-DTA transgene resulted in loss-of-function of *Rosa26* (personal communication, Kassiotis lab). Therefore, blastocysts from the experimental mating could either be *Rosa26* heterozygous (*R26*-DTA/+) or *Rosa26* homozygous (+/+), with potentially greater *Rosa26* transcriptional output. DTA expression could not be used for distinguishing the *R26*-DTA and +/+ male embryos, as DTA expression was inactive without Cre recombinase floxed-STOP excision. Therefore, homozygous versus heterozygous *Rosa26* expression was used a proxy for presence or absence of the DTA transgene. I predicted that all *R26*-DTA/+ embryos were male, as the female *R26*-DTA/+ population was embryonic lethal due to Cre-activated DTA. In the experimental +/+ blastocyst population, approximately half would be female and half male. For *Rosa26* expression in the control blastocysts, I predicted there was no grouping into females and males.

Results showed that in the experimental samples, the most lowly-expressing samples were males, and the expression was below that of all wildtype samples (bracket, Figure 10b). This population of lowly *Rosa26*-expressing males suggested that this group was the *R26*-DTA/+ group, with single copy expression of *Rosa26*. The lack of females in the

low *R26*-DTA/+ group further suggested that the *R26*-DTA/+ females have been lost prior to the blastocyst-stage, due to Cre-activated DTA expression.

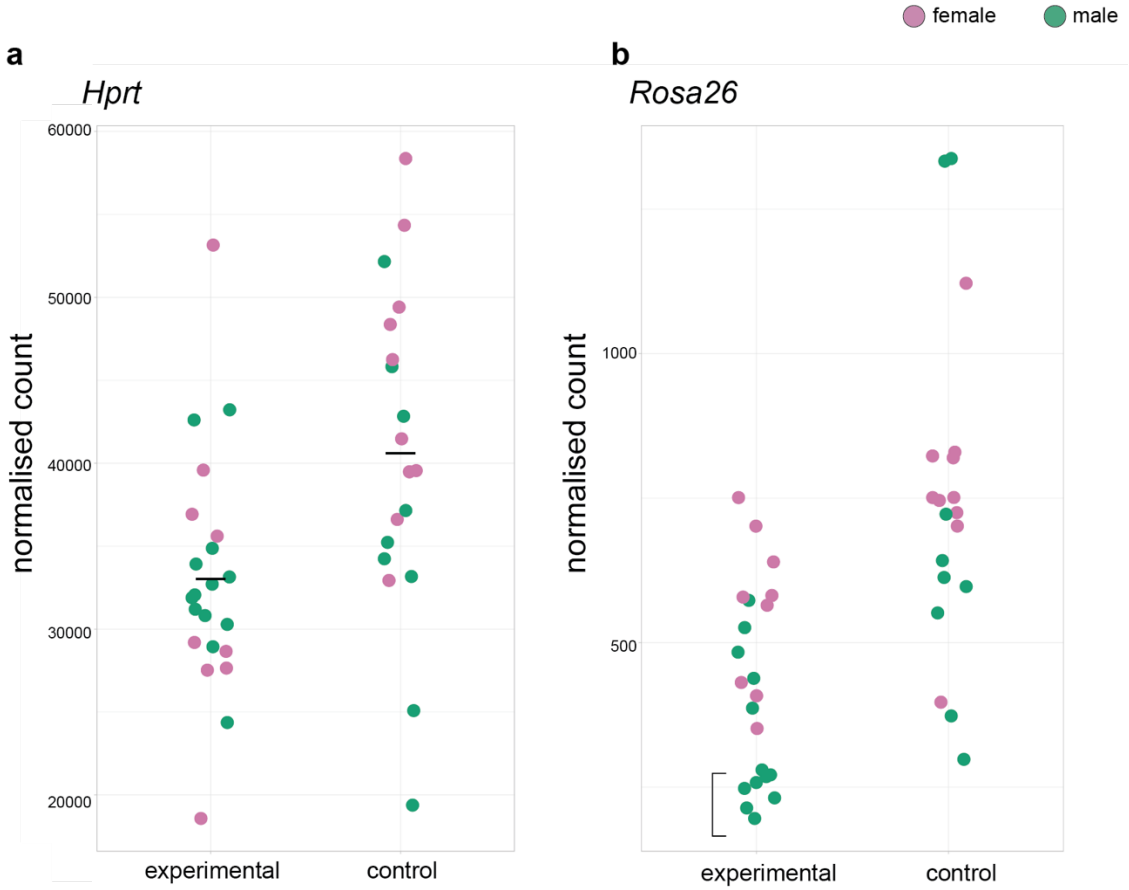

Figure 10. *Hprt* and *Rosa26* relative gene expression

Relative gene expression between experimental and control blastocysts (a) *Hprt* (b) *Rosa26.*

Conversely, the more highly *Rosa26*-expressing blastocysts consisted of males and females, suggesting this group was the +/+ population.

I next determined if there were global transcriptomic differences between the surviving blastocysts from the experimental matings and the controls. I performed a Principle Component Analysis (PCA), to determine the variance between samples. The PCA showed that there was a cluster of male samples (triangles), and a cluster of female samples (circles). There were two outliers from the two main clusters, that were both female samples (clustered in an oval outline, Figure 11a). Interestingly, these two samples corresponded to embryos collected from the same control litter and were phenotypically underdeveloped compared to the blastocyst littermates, with morphology more similar to

late-morula/early-blastocyst. Principle component 1 (PC1) accounted for 33% of the variance between samples, and therefore likely differentiated between developmental age (Figure 11a). Principle component 2 (PC2) accounted for 13% of the variance between samples, and likely differentiated between sex (Figure 11a). There was clearly no distinct clustering of experimental from control samples (Figure 11a).

Next, I generated heatmaps clustering the samples on transcriptomic similarity based on the transformed log2-fold change. I analysed the top 20 most highly differentially expressed genes across all of the samples. The two genes that showed the most highly differential expression were *Eif5a* (Eukaryotic translation initiation factor 5A) and *Hspa8* (Heat shock protein 8). However the differential gene expression of these two genes was not attributable to the condition as the upregulation of *Eif5a* and *Hspa8* was detected in experimental and control samples (Figure 11b, left first two columns). There was no hierarchical clustering of experimental or control samples based on differential gene expression.

Next, I generated a heatmap of the top 20 most highly differentially expressed genes, when quantified as differential gene expression from the average gene expression across all samples. Clustering in this way takes into account variance across all the samples. The samples were hierarchically clustered on differential gene expression, however there was no distinct hierarchical clustering between experimental versus control samples. Interestingly, the samples instead clustered mostly by sex, with four distinct gene expression profiles that were upregulated in a group of samples compared to others. These four genes were *Ddx3y, Eif2s3y, Uty* and *Xist*. The upregulation of these genes was only in male samples for *Ddx3y*, *Eif2s3y* and *Uty*, whilst *Xist* was only upregulated in females (Figure 11c). Furthermore, the two underdeveloped control samples (red box, Figure 11c) again were more hierarchically similar and clustered together, with similar patterns of gene upregulation in these two samples.

There was very little global transcriptomic differences between blastocysts from control matings, compared to blastocysts from experimental matings, based on differential gene expression analyses from RNAseq data. Therefore, I concluded that the *in utero* lethality

of X-Cre/X;*R26*-DTA/+ embryos had minimal effect on the unaffected littermates at the transcriptomic level.

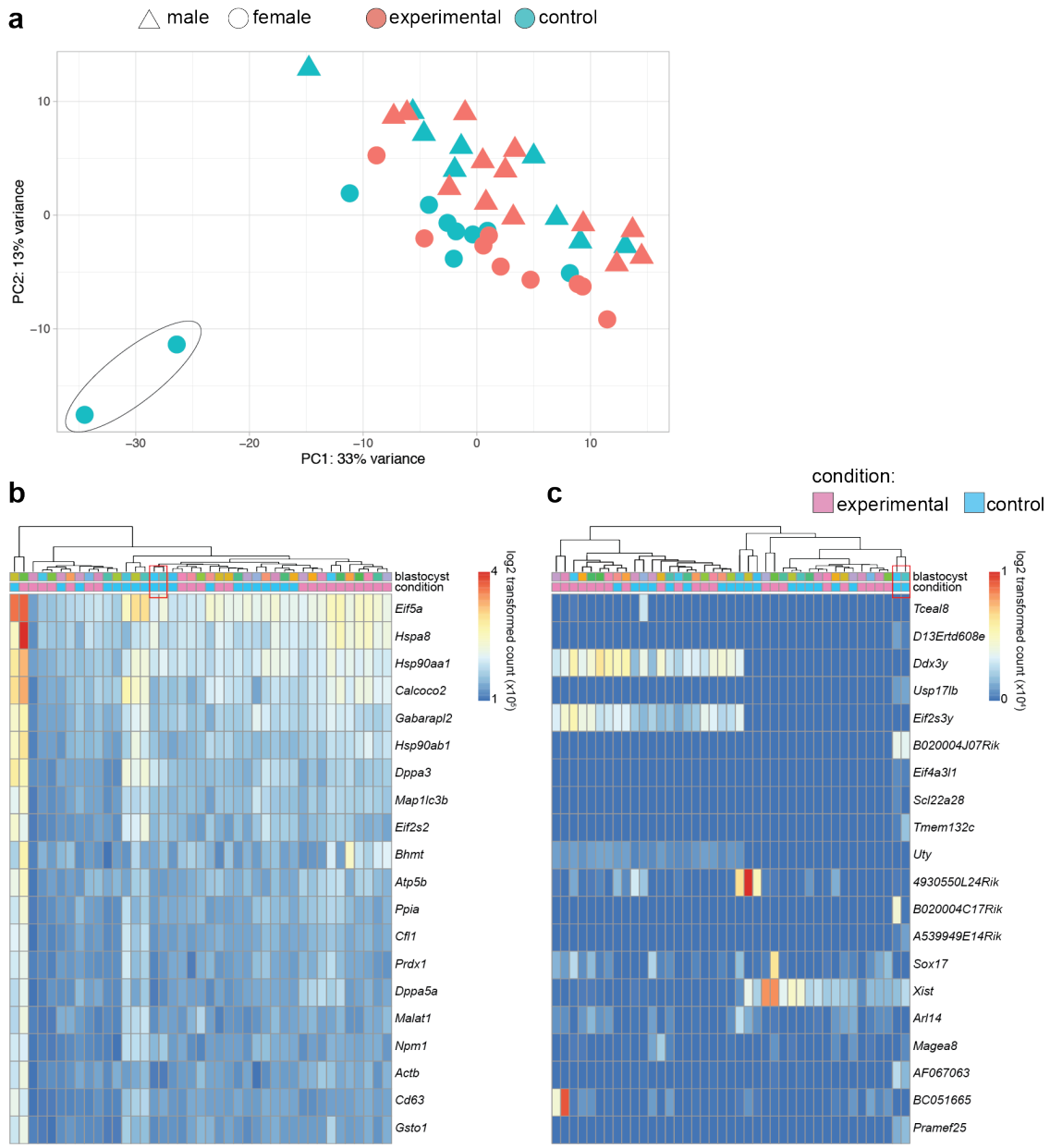

Figure 11. Differential gene expression analyses

(a) Principle component analysis on experimental versus control blastocysts. (b) Heatmap and hierarchical clustering of experimental versus control blastocysts. The top 20 most highly differentially expressed genes (log2 transformed count) are shown. (c) Heatmap and hierarchical clustering of experimental versus control blastocysts. The top 20 most highly differentially expressed genes (log2 transformed count), as differential expression from the average gene expression of all samples.

3.3 Discussion

The data presented in this chapter shows that it is possible to generate single-sex litters. Producing all-male litters was performed genetically, by X-linked Cre-induced expression of a Diphtheria Toxin A allele. Initially I was concerned that the paternal X-linked Cre allele would be transcriptionally silenced in female embryos, due to imprinted X-inactivation in the mouse pre-implantation embryo. Silencing of the Cre allele would prevent activation of the DTA allele and DTA-induced female-specific lethality. However, experiments using the Cre reporter TdTomato line showed that the X-linked Cre is active from at least E2.5, and was functional to induce Cre expression *in utero*. Although the paternal X is silenced from the 2-4 cell stage in the pre-implantation mouse embryo (Okamoto et al., 2005, Okamoto et al., 2004, Zuccotti et al., 2002), *Xist* spreads progressively *in cis* from the X inactivation centre (XIC). Therefore, given that the *Hprt* locus where the Cre transgene is integrated is not immediately adjacent to the XIC, there may have been a period of expression from X-Cre, prior to XCI.

Using the X-linked Cre, inducing DTA expression resulted in female-specific lethality. I hypothesised that the complete sex skew to all-male litters would result in the mean experimental litter size of these matings being approximately half of the mean litter size of the control matings. Although there was some reduction in mean litter size, the mean litter size was not half. This partial compensation could be because the loss *in utero* of the female offspring may allow more male embryos to implant and survive gestation through to birth. The partial compensation of litter size was also seen in later experiments, and is discussed in greater depth in 5.3.

Moreover, despite the embryonic lethality of female embryos, there appeared to be no transcriptomic detriment on the surviving embryos compared to wildtype embryos. I suspected that the X-Cre/X;*R26*-DTA/+ blastocysts were non-viable prior to the blastocyst-stage, and were therefore not harvested during blastocyst-stage collections. Therefore, the X-Cre/X;*R26*-DTA/+ embryos were not represented in the RNAseq dataset. The lack of X-Cre/X;*R26*-DTA/+ females was confirmed by an absence of lowly *Rosa26*-expressing female embryos. In humans, DTA functions by ribosylation of host

eEF-2 (elongation factor-2), disrupting eEF-2 function and inhibiting protein synthesis (Bell and Eisenberg, 1996). Therefore I predicted that if I had captured the X-Cre/X;*R26*-DTA/+ female embryos, there may be dysregulation of genes involved in protein synthesis. Although the translation initiator gene *Eif5a* appears to be upregulated in some samples, the upregulation did not appear to be unique to the experimental samples. Therefore it is unlikely this upregulation is a result of embryonic littermate lethality. Furthermore, there was no differences in relative *Eef-2* gene expression between experimental and control blastocysts in this dataset. It is likely that any transcriptomic differences seen at the individual level is due to *in utero* variability, for example differences in developmental time, and/or lysis, cDNA and library preparation procedure. The finding that the littermates are unaffected at the transcriptomic level is important if a similar strategy is to be undertaken in agricultural species, as the required litter-mates must develop similarly to wildtype.

These experiments generated all-male litters. However, using an inducible DTA transgene to cause female lethality may not be immediately translatable to agriculture and livestock. Most commonly in agriculture it is the females that are required; for example in dairy cow or layer hen production. In the future it may be possible to reverse this technology by inserting a Cre recombinase allele onto the Y chromosome. In this strategy, the Cre recombinase would be uniquely inherited by the sons, resulting in a sex skew towards female offspring.

The main caveat of the inducible DTA mating strategy is that the live offspring carry an allele encoding a toxin. Although theoretically DTA expression is only induced upon recombination by Cre (Marques et al., 2009), it is possible that DTA expression could be leaky. If this technology was to be translated to an agricultural application, the public may be apprehensive of livestock or farm-produce encoding a toxin. The problem of surviving offspring carry an allele encoding a toxin could be overcome by using a CRISPR-Cas9 bi-component system, where the Cas9 allele is lost with the unrequired sex. In this approach, a known housekeeping gene with essential embryonic function can be targeted for loss-of-function knock-out. Furthermore, the embryos could be collected at multiple developmental time points, to assess the CRISPR-Cas9 induced mutation

dynamics at the target locus. Furthermore, in the CRISPR-Cas9 approach, the Cas9 and sgRNA can be genetically segregated to control inheritance of both components. The sgRNA would be inherited by the surviving offspring. This sgRNA allele is not a protein-coding gene, and therefore is clearly advantageous compared to a toxin protein-encoding gene. The development of the CRISPR-Cas9 approach will be discussed in the next three Chapters.

Chapter 4. Results 2: Screening highly mutagenic sgRNAs to generate an sgRNA expressing knock-in mouse line

4.1 Introduction

CRISPR-Cas9 genome editing requires two essential components; a Cas9 endonuclease and a single guide RNA (sgRNA). The sgRNA is made up of two regions. The first region is 20 nucleotides long and is complementary to a target site, adjacent to a PAM. The second region of the sgRNA guides the Cas9 to the target site to induce DNA DSBs. The DNA DSB is repaired either by non-homologous end joining (NHEJ) or homology directed repair (HDR), discussed in more detail in 1.5.1.1 and 1.5.1.2, respectively. NHEJ is error-prone, and results in indel mutations at the target site. Frame-shift mutations at the target site can disrupt the reading-frame of the gene, resulting in loss-of-function.

To generate single-sex litters, the two components of the CRISPR-Cas9 system have to be genetically segregated in a "bi-component" system (Figure 12). In this strategy, a Cas9 transgene, is integrated onto the X- or Y chromosome and carried by the father (similarly to the X-Cre transgene described in Chapter 3). The Cas9 transgene is uniquely inherited by daughters (X-Cas9) or sons (Y-Cas9) dependent on which sex chromosome the Cas9 is integrated. The sgRNA targets a housekeeping gene with an essential role in early embryonic development. This "lethal sgRNA" is carried bi-allelically by the mother and therefore inherited by all offspring irrespective of sex. Only when both components of the CRISPR-Cas9 system are inherited would mutations at the target housekeeping gene occur, resulting in embryonic non-viability.

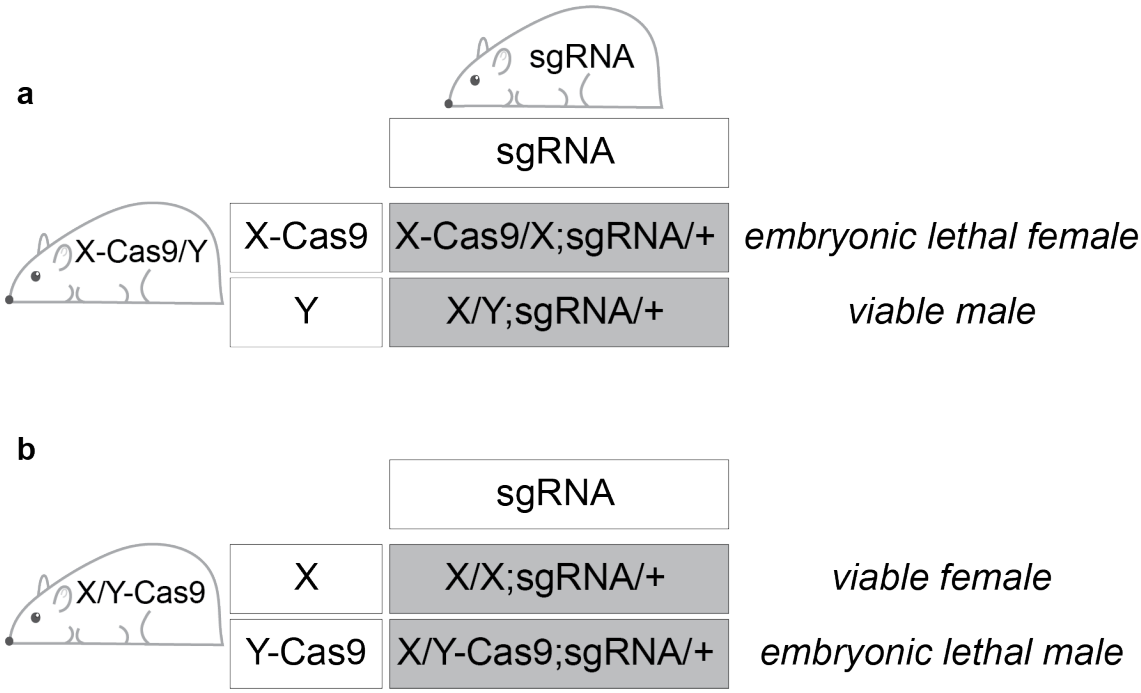

Figure 12. CRISPR-Cas9 bi-component system for sex-specific lethality

(a) Female-specific embryonic lethality, producing an all-male litter. An X-Cas9/Y male is crossed with a homozygous sgRNA-expressing female. Inheritance of both the X-Cas9 and sgRNA alleles in females induces mutations at an essential housekeeping gene, resulting in loss-of-function and embryonic non-viability. (b) Male-specific embryonic lethality, producing an all-female litter. An X/Y-Cas9 male is crossed with a homozygous sgRNA-expressing female. Inheritance of the Y-Cas9 and sgRNA allele in males induces mutations in an essential housekeeping gene, resulting in loss-of-function and embryonic non-viability.

Inducing embryonic lethality during pre-implantation embryonic development would maximise the length of time possible to induce a sex ratio skew. Ideally, the Cas9 and sgRNA components would be under constitutive promoters, to ensure onset of expression soon after EGA in the mouse embryo. Targeting an essential housekeeping gene for knock-out requires a highly efficiency sgRNA to eliminate the protein-coding function of the gene. Ideally the CRISPR-Cas9 would induce a frame-shift mutation to disrupt the reading-frame of the gene, producing a loss-of-function phenotype.

To translate the bi-component strategy *in vivo*, the final aim was to generate a transgenic mouse line expressing a highly efficient sgRNA expressed from an autosome. However standard *Streptococcus pyogenes* CRISPR-Cas9 HDR approaches could not be used to generate this sgRNA-knock in mouse model. Contemporaneous expression of the sgRNA transgene and residual Cas9 expression may induce mutations in the essential

housekeeping gene and embryonic lethality. Thereby, any positive knock-in embryos would be lost. I looked to using different modes of generating sgRNA transgenic mouse lines. One method of generating targeted knock-ins is by TARGATT technology (Zhu et al., 2014, Tasic et al., 2011, Chen-Tsai, 2019, Rossant et al., 2011). Tasic *et al* (2011), produced a transgenic mouse model containing a tandem of three attP sites (attPx3) inserted into the permissive autosomal *Hipp11* locus (*H11*; Hippenmeyer et al., 2010) and the *Rosa26* locus (*R26*; Friedrich and Soriano, 1991, Tasic et al., 2011). Repair templates containing attB sites, are integrated into the attPx3 locus by C31ϕ integrase-mediated recombination. Due to the transgene insertion being generated using non-CRISPR based approaches, I hypothesised that expression of the lethal sgRNA, either transiently expressed or stably integrated as a transgene, would have no lethality effect on the developing embryo.

4.2 Results

4.2.1 Deriving *Rosa26* Cas9 eGFP embryonic stem cells

To test multiple sgRNAs *in vitro* whereby the Cas9 and sgRNA are genetically segregated, I required an *in vitro* culture of Cas9-expressing cells. I chose to derive embryonic stem cells (ESCs) from the Gt(*ROSA*)26Sor$^{tm1.1(CAG-Cas9*;eGFP)Fezh}$/J (hereafter referred to as "*R26*-Cas9") pre-existing mouse line (Platt et al., 2014) for this purpose. This mouse strain constitutively expresses a SpCas9 (hereon referred to "Cas9") endonuclease driven by a constitutively active CAG promoter. Furthermore, an eGFP fluorescent reporter is linked to the Cas9 via a 2A polypeptide. eGFP expression was used as a proxy for Cas9 expression in downstream experiments.

To derive the ESC line, homozygous *R26*-Cas9 males were set up for timed mating with wildtype (MF1) females (+/+). I collected E3.5 embryos and derived three ESC lines (#1, 2 and 3) in 2i+LIF conditions (see 2.1.3 and 2.6.2, Figure 13a). ESCs derived in 2i+LIF conditions are able to retain euploidy and they express the pluripotent markers of the epiblast (Mulas et al., 2019a). Female (XX) 2i+LIF derived ESCs are more susceptible to hypomethylation (Zvetkova et al., 2005, Choi et al., 2017), and loss of an X chromosome (Zvetkova et al., 2005) compared with male ESCs, reducing their ability to germline transmit in chimeras. Although I was not aiming to generate chimeras from these ESCs, I nevertheless aimed to do all downstream experiments with male (XY) ESCs. I determined the sex of each of the ESC lines by PCR amplification of the *Sly* gene, a Y chromosome-linked gene (Figure 13b). In two of the three derived ESC lines (#2 and #3) I could amplify *Sly,* indicating the presence of a Y chromosome. An autosomal gene *Myogenin* was also amplified as a PCR control (Figure 13b). One male ESC line (#2) was carried forwards for future experiments.

Next, I confirmed eGFP and Cas9 expression from the *R26*-Cas9 ESCs. Theoretically, all derived ESC lines should have been *R26*-Cas9 hemizygous, due to using a homozygous father for timed matings. I performed reverse transcriptase PCR (RT-PCR) with minus-

RT controls for gDNA contamination, and a β-actin housekeeping gene control. *R26*-Cas9 and XY wildtype ESC lines were positive for β-actin expression whilst the *R26*-Cas9 ESCs expressed eGFP and Cas9 (Figure 13c). Furthermore, I confirmed that eGFP expression could be visualised by fluorescence microscopy in male ESC line #2 (Figure 13d).

I also performed quantitative-PCR (qPCR) to confirm expression of Y chromosome genes *Uty* and *Eif2s3y* to ensure the *R26*-Cas9 #2 ESCs were male (Figure 13e). In all qPCR experiments, *R26*-Cas9 XY and XX gene expression was normalised to wildtype *Gapdh* expression. Results showed that the *R26*-Cas9 ESCs expressed the Y chromosome genes. I compared expression with *R26*-Cas9 line #1, which was female, and did not express *Uty*/*Eif2s3y*. I also assessed expression of *Pou5f1,* a highly-expressed ESC pluripotency marker (Nichols et al., 1998) as a qPCR control. *Pou5f1* was expressed in all samples. Lastly, I confirmed that eGFP was expressed in the *R26*-Cas9 male #2 and female #1 ESCs but not in the wildtype control (Figure 13e).

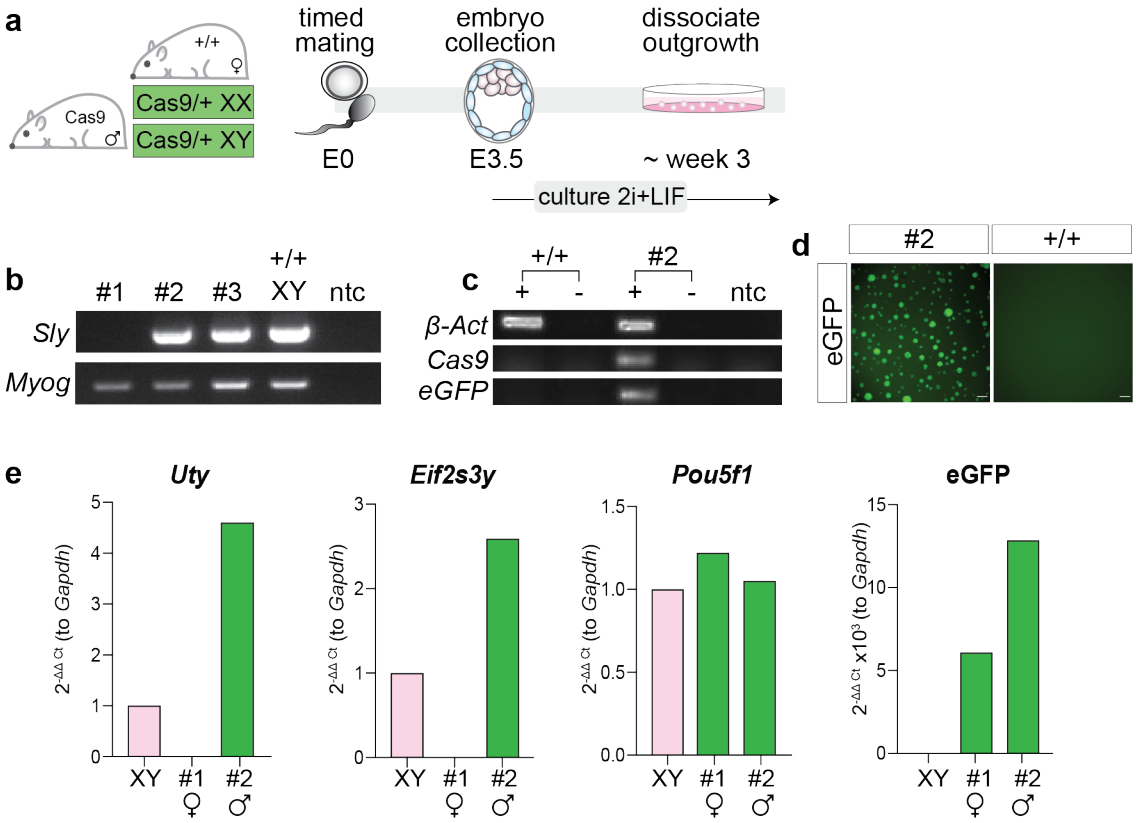

Figure 13. Derivation of *R26*-Cas9 embryonic stem cells

(a) A male homozygous *R26*-Cas9 ("Cas9") was mated with wildtype MF1 females (+/+) and embryos collected at E3.5. Embryos were plated in 2i+LIF ESC medium until expanded. The outgrowth was dissociated into ESCs, plated and maintained in 2i+LIF. (b) Three ESC lines (#1-3) were genotyped by PCR for *Sly* (Y chromosome) and for *Myogenin* (autosome) including a water no template control (ntc). (c) RT-PCR to confirm expression of Cas9/eGFP in male *R26*-Cas9 ESCs (#2) plus minus-RT and ntc controls. (d) Fluorescence microscopy of ESC line #2 Scale bar=50µm (e) qPCR analysis to confirm expression of Y chromosome genes *Uty* and *Eif2s3y*, ESC marker *Pou5f1* and the eGFP transgene.

Overall, these results showed successful derivation of male *R26*-Cas9 ESCs with detectable Cas9 and eGFP expression. Male ESC line #2 was carried forward to facilitate low-throughput screening of multiple sgRNAs. In all future experiments using this line #2, the ESCs are referred to as "*R26*-Cas9".

4.2.2 Candidate embryonic lethal gene: *Topoisomerase 1*

I assessed candidate genes that when knocked-out, would induce embryonic non-viability at pre-implantation (http://www.informatics.jax.org/mp/annotations/MP:0011094). I focused on genes that showed a complete penetrance of lethality, and had a role in DNA

repair and replication, as knock-out of this gene is likely to have a function across many cell types during embryonic development.

I chose to target the essential housekeeping gene *Topoisomerase 1* (*Top1*), a protein-coding gene present on mouse chromosome 2. Mouse *Top1* has 21 exons, encoding a protein with four domains. The C-terminal exons 13 to 20 encode the essential DNA-binding domain and catalytic site. Exons 11 and 12 encode the coiled-coil linker domain. *Top1* also encodes two poorly conserved N-terminal low-complexity domains, exons 3-4 and 7-8 (Figure 14a; Wright et al., 2015).

TOP1 is a monomeric protein that functions by encircling the dsDNA helix to form a complex. Once the complex is formed, a single tyrosine within an active site functions to induce a 3'-phosphotyrosyl intermediate in a nicked single strand of DNA, allowing the supercoils to relax when the DNA rotates, and eventually re-ligating the nicked DNA strand (Pommier et al., 2010). TOP1 protein is highly conserved, particularly at the C-terminal domain. The amino-acid alignment between four species, mouse, cow, chicken and rat is shown in Figure 14b with the DNA-binding domain and catalytic site shown by the positioning of the grey box below the alignment (Figure 14b).

Studies have taken advantage of a LoxP-Cre system, generating conditional knock-outs, driven by a cell type specific reporter to investigate the precise function of *Top1*. Mabb *et al*, generated a conditional knock-out mouse line by inserting two LoxP sites flanking exon 3 (Mabb et al., 2016). Previous studies have also taken advantage of the highly conserved C-terminal region of the *Top1* gene, generating *Top1* null mutant mouse ESCs by replacement of exon 15 with a neomycin cassette (Morham et al., 1996) or by inserting two LoxP sites flanking exon 15 to generate a conditional knock-out (Kobayashi et al., 2011). In the mouse, homozygous knock-out of *Top1* results in embryonic non-viability, occurring around the 4-16 cell stage of pre-implantation development (Morham et al., 1996). In cultured mammalian cells, homozygous *Top1* knock-out is lethal (Morham et al., 1996) and TOP1 reduction causes genetic instability (Kobayashi et al., 2011), chromosome breaks and translocations (Tuduri et al., 2009). Therefore, it is essential *Top1* expression is tightly controlled, to ensure genomic stability (Wang, 2002).

Figure 14. *Top1*/TOP1

(a) The mouse *Top1* gene with exons in yellow. (b) TOP1 is highly conserved among multiple species. The amino acid level alignment between cow, chicken, mouse and rat. The grey box indicates the amino acid sequence coding for the essential DNA binding and catalytic domain.

The pre-implantation embryonic lethal phenotype of *Top1* suggested that this gene would be appropriate to target in the bi-component system. I aimed to develop a method to low-throughput screen a number of sgRNAs targeting different exons of the *Top1* locus, and determine the mutation efficiency of each of the sgRNAs.

4.2.3 Generating the lethal sgRNA plasmid to screen sgRNAs

For proof-of-principle experiments generating single-sex litters by CRISPR-Cas9 induced knock-out, *Top1* was selected as the candidate gene. *Top1* exons 13 to 20 encode the highly-conserved catalytic core and DNA-binding domain (Wright et al., 2015, Morham et al., 1996, Kobayashi et al., 2011). Therefore, I designed one sgRNA targeting exon 15 and one sgRNA targeting exon 16 (Figure 15a). Previous CRISPR-Cas9 knock-out strategies have designed sgRNAs targeting proximally to the ATG start codon, to disrupt the reading frame early (Doench et al., 2016). I designed a third sgRNA targeting close to the ATG codon, in exon 2 (Figure 15a).

I utilised the pX333 plasmid vector (Maddalo et al., 2014) as a backbone for sgRNA cloning. I exchanged the Cas9 cassette for an mCherry reporter driven by a constitutive Cbh promoter (Figure 15b). Expression of the sgRNA and mCherry reporter could not be driven by the same promoter and cleaved via a 2A polypeptide, because the sgRNA does not encode a protein. Therefore sgRNA expression is driven separately by a human U6 promoter. The pX333 plasmid was advantageous because it contains a tandem of sgRNA cassettes under separate U6 promoters The two sgRNA scaffolds contain different restriction enzyme directional cloning sites (BbsI and BsaI). Therefore, two sgRNAs can be expressed from the same plasmid. The BbsI restriction enzyme site was also contained within the mCherry gene sequence, and this BbsI site was edited to contain a silent mutation (GAAGAC to GAAAAC) to prevent sgRNA cloning disrupting the mCherry coding sequence. The final sgRNA plasmid (annotated as “pLethal”, Figure 15b) contained the Cbh-mCherry reporter and two U6-sgRNA scaffolds, and was transfected in HEK293T cells. The mCherry expression was detectable by fluorescence microscopy (Figure 15c).

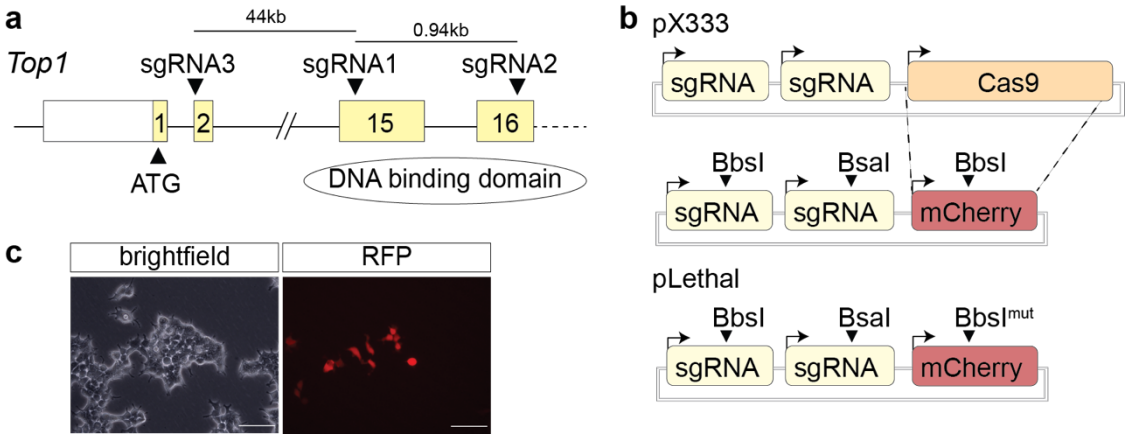

Figure 15. Generating the "pLethal" plasmid for screening sgRNAs

(a) The *Topoisomerase 1* (*Top1*) locus highlighting the target regions of each sgRNA; sgRNA1 (exon 15) , 2 (exon 16) and 3 (exon 2) (b) Generation of the pLethal plasmid by modification of pX333; replacement of the Cas9 cassette with the mCherry reporter. The plasmid also contains two tandem U6 promoters for separate sgRNA expression. A silent mutation was introduced into a BbsI restriction enzyme site within the mCherry sequence (c) Confirmation of mCherry expression by transient transfection in HEK293T cells. Scale bars=100µm.

The individual *Top1* sgRNAs (1, 2 and 3) were cloned into separate pLethal plasmids, to be driven by the first U6 promoter, and confirmed by Sanger sequencing. I also cloned an sgRNA targeting the Y chromosome gene *Sry* (targeting the single exon 1) into a pLethal plasmid. This *Sry* sgRNA could be used as a non-lethal control sgRNA. This sgRNA has been used previously in the Turner lab, to generate *Sry* mutations at high efficiency (unpublished). After generating the pLethal plasmids containing *Top1* and *Sry* sgRNAs, I next aimed to evaluate the mutation efficiency of the lethal *Top1* sgRNAs *in vitro*.

4.2.4 Generating mutations at *Top1* and evaluating mutation efficiency

In order to assess mutagenic efficiency of the three *Top1* sgRNAs *in vitro*, I designed a strategy whereby the pLethal plasmids could be individually transfected into *R26*-Cas9 ESCs and the transfected and un-transfected ESCs sorted into separate populations by fluorescence activated cell sorting (FACS). The ESCs already express eGFP and Cas9, and pLethal-transfected ESCs would also transiently express mCherry and the sgRNA.

Therefore, two populations would be sorted; mCherry+eGFP+ and eGFP+ only. This 'traffic-light' system of mCherry and eGFP reporter expression allowed me to use the fluorescence as a read-out for presence of sgRNA and Cas9, respectively. The "single-positive" eGFP+ ESCs, which express Cas9 in isolation, functioned as an internal control to determine the basal variability at a target locus. Whilst in the mCherry+eGFP+ "double-positive" cell populations, both components of the CRISPR-Cas9 system are present, allowing mutations at the target exon.

My first aim was to determine the optimum time point post-transfection in which to perform FACS. Knock-out of *Top1* in ESCs is lethal (Morham et al., 1996) therefore the ESCs would need to be sorted by FACS before ESC death, but after sufficient pLethal-transfection and expression to capture mCherry+ ESCs. To assess the transfection efficiency, I transiently transfected *R26*-Cas9 ESCs with the pLethal plasmid containing the non-lethal *Sry* sgRNA. In four separate plates, $1x10^6$ *R26*-Cas9 ESCs were transfected and FACS performed 4, 24, 48 and 72 hours post-transfection. There was an increase in mCherry+eGFP+ ESCs with increasing time post transfection (Figure 16a). This was quantified as a percentage of mCherry+eGFP+ ESCs from the total amount of sorted ESCs. I decided on a 48 hour post-transfection time point for FACS for all future *Top1* sgRNA experiments, a compromise between high rate of transfection, and minimum time post-transfection to minimalize ESC death.

My next aim was to sort ESCs post-transfection with the *Top1* sgRNAs, to assess the mutation efficiency of each sgRNA. Each sgRNA-containing pLethal plasmid was transfected into a separate plate of *R26*-Cas9 ESCs ($1x10^6$ ESCs per transfection). After 48 hours the transfected cells were sorted by FACS (representative FACS plots shown in Figure 16c). The FACS was performed on DAPI negative ESCs, i.e. live ESCs, for both the mCherry+eGFP+ and eGFP+ populations. The mCherry+eGFP+ ESCs were gated above 10^3 fluorescence units for mCherry (Figure 16c). The eGFP+ ESCs was gated at approximately $10^{3.5}$ to $10^{4.5}$ fluorescence units for GFP (Figure 16c). I kept the mCherry and GFP gates discrete, as not to capture contaminating ESCs between populations. The two populations, mCherry+eGFP+ or GFP+ were sorted into separate tubes and the ESCs lysed (Figure 16b). After lysis, the ESC extract was PCR amplified at the target exon

regions for each sgRNA. The PCR amplicons were then deep-sequenced by MiSeq and analysed by bioinformatic analysis to evaluate the range of reads at the target site (Figure 16b) using a previously published R package, CrispRVariants (Lindsay et al., 2016), (see 2.9.2). A bioinformatic approach to analyse the spectrum of reads ensured that I could capture all variability of mutations at each of the target loci. The mutation efficiency output of the R package was calculated by evaluating the number of reads containing an insertion or deletion (indel) mutation, as a percentage of the total number of reads. The sgRNAs were tested in triplicate, and the mean mutation efficiency was calculated. Single nucleotide variants (SNVs) were not counted as mutations.

The results showed that transfection of sgRNA1, targeting *Top1* exon 15, produced the highest mutation efficiency (Figure 16d). The mean mutation efficiency in mCherry+eGFP+ ESCs post-transfection with sgRNA1 was 52.2%. The mutation rate was comparably greater than the mean mutation efficiency of 22.0% and 28.9% after transfection of sgRNA2 or 3, respectively (Figure 16d). In the eGFP+ populations, the mean mutation efficiency was 1.15% (sgRNA1), 0.43% (sgRNA2) and 1.18% (sgRNA3; Figure 16d). I attribute the low but detectable non-Cas9-induced mutation rate in the eGFP+ population to PCR error, endogenous variability at *Top1,* and/or contamination of mCherry+eGFP+ ESCs in the eGFP+ population. These explanations will be elaborated on further in 4.3.

I then determined if there was a reduction in TOP1 protein after transfection with each of the single sgRNAs. The pLethal plasmids were transfected as previously, and the mCherry+eGFP+ and eGFP+ ESC populations sorted by FACS 48 hours post-transfection. After sorting, bulk protein was extracted and western blot performed for TOP1 using either an N-terminal or C-terminal antibody. In the eGFP+ non-transfected ESCs TOP1 was expressed, when using either the N- or C-terminal antibody (Figure 16e). Conversely in the mCherry+eGFP+ ESCs there was a complete loss of TOP1 expression, whether using either the N- or C-terminal antibody (Figure 16e).

Successful transfection of a single sgRNA into *R26*-Cas9 ESCs induced mutations at the target exons *in vitro*. However, using multiple sgRNAs targeting a gene of interest may

increase the probability of generating a null mutation. Therefore, given the highly mutagenic activity of *Top1* sgRNA1, I cloned a second *Top1* sgRNA into the sgRNA1 plasmid, driven by the second U6 promoter. Two new pLethal plasmids were generated: sgRNA1+2, targeting exon 15 and 16, and sgRNA1+3, targeting exon 15 and 2. After generating the 'double sgRNA' pLethal plasmids, I aimed to determine whether the second U6 promoter would work equivalently to the first to drive sgRNA expression. The single sgRNA1 plasmid, and two double sgRNA plasmids (1+2 and 1+3) were transfected into individual plates of *R26*-Cas9 ESCs, as earlier described. The mCherry+eGFP+ and eGFP+ ESCs were sorted by FACS into separate populations. When the double sgRNA plasmids were transfected, two separate PCR MiSeq products were amplified for each of the individual target sites for each sgRNA. The two target exons for the double sgRNA plasmids could then be analysed separately to assess mutation efficiency.

After amplification of the expected target sites and sequencing by MiSeq, the mutation efficiency was determined using the CrispRVariants pipeline. The mean mutation efficiency of the single sgRNA1 was 63.9% in mCherry+eGFP+ ESCs, compared with 1.1% in eGFP+ ESCs (Figure 16f). Post-transfection with sgRNA1+2, the mean mutation efficiency for mCherry+eGFP+ ESCs at exon 15 (sgRNA1) was 64.5%. At this exon, the eGFP+ only ESCs had a mean mutation efficiency of 1.5%. When the second sgRNA was sgRNA2, the induced mutations are at exon 16, with a mean mutation efficiency of 50.8%, compared with 0.3% in the eGFP+ population (Figure 16f). After transfection with plasmid sgRNA1+3, in the mCherry+eGFP+ population, sgRNA1 generated a mean mutation efficiency of 59.1%. The sgRNA3 generated a mean mutation efficiency of 61.3% at exon 2. In the eGFP+ only population, exon 15 (sgRNA1) showed a mutation efficiency of 1.3%, whilst at exon 2 (sgRNA3) the mutation efficiency was 0.7% (Figure 16f). This data indicated that in a double sgRNA plasmid, the second sgRNA induces mutations at the target exon and that the tandem U6 promoter can drive sgRNA expression. However having additional sgRNAs expressed from a single plasmid, does not increase the efficiency of indel mutations generated at individual exons.

Introducing two sgRNAs (and Cas9) with close target proximity can result in complex genomic rearrangements between the two sgRNA sites (Kosicki et al., 2018, Allen et al.,

2018); Figure 16h). I amplified a PCR product from mCherry+eGFP ESCs, using primers aligning up- and downstream of sgRNAs 1 and 2, which are approximately 940bp apart (Figure 16g). Two products were amplified, the expected 1.1kb product and one at approximately 230bp (Figure 16g, arrow heads). I cloned the PCR products into a cloning vector plasmid and Sanger sequenced the clones. I firstly investigated the mutation efficiency at each of the two DSB sites individually. Results showed that in 28% colonies sequenced, there was a single mutation in exon 15 (sgRNA1). In 6% of colonies sequenced, there was a single mutation at exon 16 (sgRNA2). In 50% of colonies, there was a mutation at both exons. Only 16% products presented wildtype sequence at both sgRNA1 and sgRNA2 cut sites (Figure 16g). This data shows that Cas9-induced DSBs at two loci do not always result in indels simultaneously at one dsDNA molecule.

Secondly, I investigated if there was a large deletion between the sgRNA1 and 2 target sites. Two clones contained a ~250bp product that aligned to *Top1* DNA sequence 5' to sgRNA1 and 3' to sgRNA2, with none of the intervening DNA sequence (Figure 16h). This chimeric PCR product confirmed that transfection with a double sgRNA plasmid can induce large deletions of the intervening sequence between two sgRNA target sites. Unfortunately, between exon 2 (sgRNA3) and exon 15 (sgRNA1), there is over 44kb of sequence, preventing any PCR amplification flanking these two cut sites.

Given the known mutagenic potential of the single sgRNA1 plasmid, and the two double sgRNA plasmids; sgRNA 1+2 and sgRNA 1+3, I then lastly aimed to investigate the mutagenic efficiency of these three sgRNA plasmids over time. For each sgRNA-containing plasmid, four separate plates of *R26*-Cas9 ESCs were transfected. The ESCs were sorted into mCherry+eGFP+ and eGFP+ only populations by FACS at 4, 24, 48 and 72 hours post-transfection, in duplicate, as earlier described. The populations of mCherry+eGFP+ and eGFP+ only ESCs could be sorted by FACS as early as 4 hours post-transfection by reporter expression. In all three transfections with each pLethal plasmid, mutations occurred 4 hours post-transfection. The highest mutation efficiency (46.2%) was present at exon 16, induced by sgRNA2 in the double sgRNA plasmid (sgRNA1+2). At 24 hours, the most highly efficient sgRNAs were sgRNA1 (single sgRNA, 65.9%) and sgRNA3 (71.2%) expressed from the sgRNA1+3 plasmid (Figure

16i). At 48 hours post-transfection, the time point used for all earlier experiments, the most efficient sgRNA was sgRNA2 (in sgRNA1+2, 69.9%) and sgRNA1 alone (60.7%). This result is echoed at 72 hours (Figure 16i).

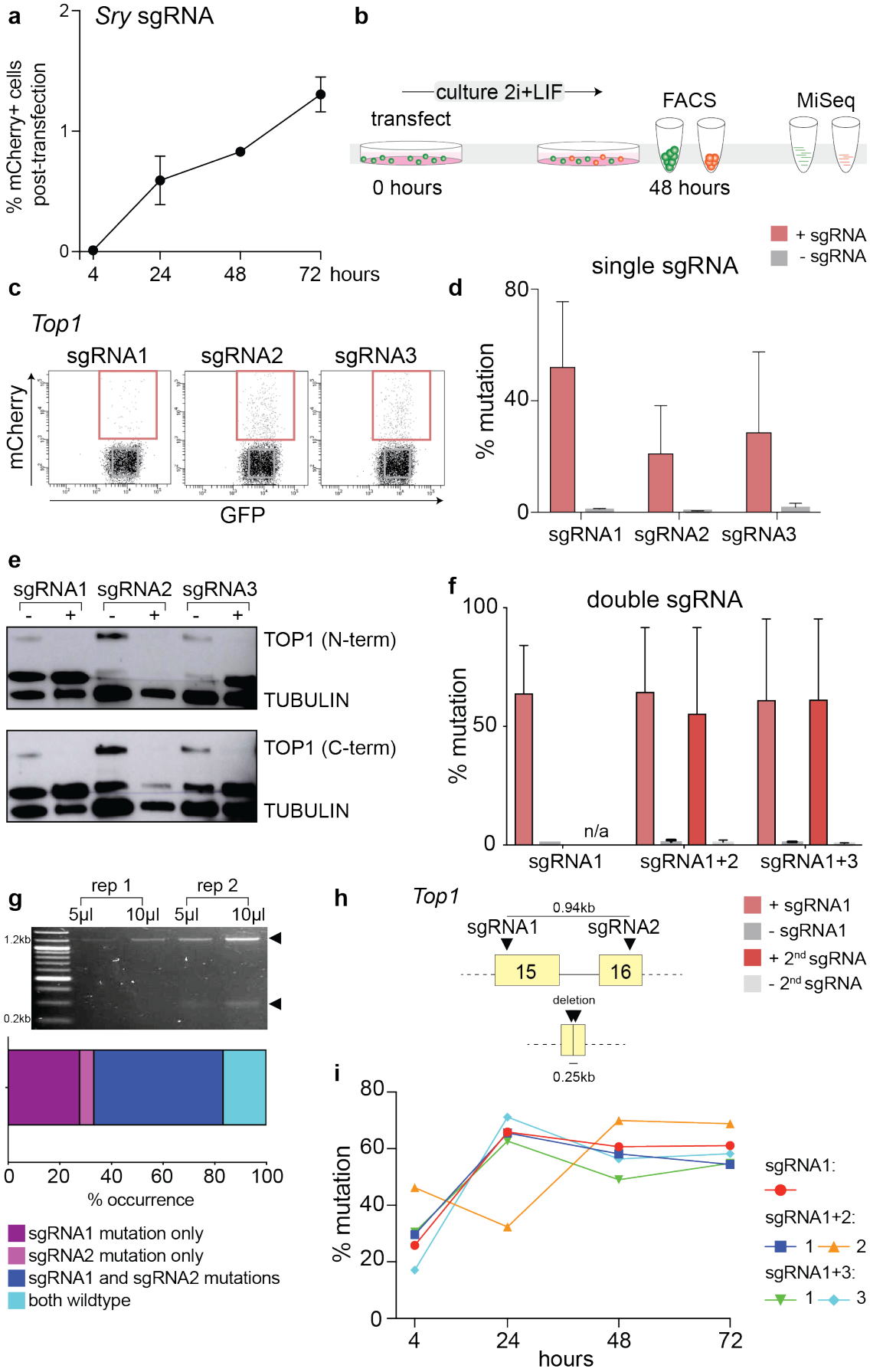
a
Sry sgRNA
% mCherry+ cells post-transfection
2
1
0
4
24
48
72 hours
b
culture 2i+LIF
transfect
FACS
MiSeq
0 hours
48 hours
+ sgRNA
- sgRNA
c
Top1
sgRNA1
sgRNA2
sgRNA3
mCherry
GFP
d
single sgRNA
% mutation
80
40
0
sgRNA1
sgRNA2
sgRNA3
e
sgRNA1
sgRNA2
sgRNA3
- + - + - +
TOP1 (N-term)
TUBULIN
TOP1 (C-term)
TUBULIN
f
double sgRNA
% mutation
100
50
0
n/a
sgRNA1
sgRNA1+2
sgRNA1+3
g
rep 1
rep 2
5µl
10µl
5µl
10µl
1.2kb
0.2kb
0
20
40
60
80
100
% occurrence
sgRNA1 mutation only
sgRNA2 mutation only
sgRNA1 and sgRNA2 mutations
both wildtype
h
Top1
0.94kb
sgRNA1
sgRNA2
15
16
deletion
0.25kb
+ sgRNA1
- sgRNA1
+ 2nd sgRNA
- 2nd sgRNA
i
% mutation
80
60
40
20
0
4
24
48
72
hours
sgRNA1:
sgRNA1+2:
1
2
sgRNA1+3:
1
3

Figure 16. Low-throughput sgRNA screen

(a) Assessment of the percentage (transfection efficiency by FACS) of mCherry+eGFP+ ESCs using a non-lethal sgRNA targeting Y chromosome gene *Sry*. (b) Strategy for transfection of *R26*-Cas9 ESCs with different sgRNAs (with mCherry reporter), followed by FACS of ESCs 48 hours post-transfection and MiSeq analysis to assess mutation efficiency in different ESC populations. The sorting by FACS was performed by the Flow Cytometry STP. (c) Representative FACS plots of *Top1* sgRNA1, 2 and 3, 48 hours post-transfection. (d) Mutation efficiency of sgRNA1, 2 and 3 at the expected target exon in *Top1*. The mutation efficiency is quantified by MiSeq of a PCR amplicon of the target region, followed by analyses using the CrispRVariants R package pipeline (Lindsay et al, 2016). (e) Western blot of TOP1 using an N-terminal and C-terminal antibody in mCherry+eGFP+ ESCs (+sgRNA) and eGFP+ only (-sgRNA) ESCs for each sgRNA. The western blot was performed by Valdone Maciulyte (Turner lab). (f) Mutation efficiency when two sgRNAs are inserted into the sgRNA1 pLethal plasmid for each of the target exons. (g) Assessment of occurrence of two separate indel mutations at exon 15 and exon 16 by sgRNA1 and 2 by cloning and Sanger sequencing. (h) Deletion of the intervening DNA sequence when using a double sgRNA approach; sgRNA1+2 and production of a chimeric PCR product. (i) Time course experiment quantifying the mutation efficiency at *Top1* for sgRNA1, 2 and 3 after FACS; 4, 24, 48 and 72 hours post-transfection. All of the MiSeq library preparations were performed by the GeMS STP. All of the MiSeq sequencing was performed by the ASF STP. The CrispRVariants R package pipeline was performed by Jasmin Zohren (Turner lab).

Overall, this data highlighted the potential of inducing mutations at the *Top1* locus using a bi-component CRISPR-Cas9 system *in vitro*. Results showed that sgRNA1 gave the highest efficiency of mutation when expressed as a single sgRNA, and induced mutations as rapidly as 4 hours post-transfection. Although sgRNA2 and 3 gave considerable mutation efficiency, particularly in combination with sgRNA1 in the double sgRNA plasmids, there was a complexity in analysing the mutation types when large intervening sequence deletions arose. Therefore, given this highly mutagenic potential of sgRNA1 as a single sgRNA, combined with the simplicity of screening for single mutations at a single target locus, I next wished to investigate the dynamics and spectrum of mutations induced by sgRNA1.

4.2.5 Dynamics of sgRNA 1 CRISPR-Cas9 induced mutations

Top1 loss-of-function is more likely if the mutation is a frame-shift. Therefore, I aimed to assess the spectrum of *Top1* exon 15 mutations in mCherry+eGFP+ ESCs, after transfection with sgRNA1, the sgRNA that created the highest mutation efficiency.

Firstly I performed an *in silico* prediction of the mutational outcome of sgRNA1, using a published tool, "CRISPOR"(crispor.tefor.net, (Concordet and Haeussler, 2018, Haeussler et al., 2016). This tool produces an "out-of-frame" score, based on a prediction of the number of clones that would contain frame-shift indel mutations at the target site, scored out of 100 (Bae et al., 2014). The out-of-frame score for sgRNA1 was 72, a high probability of generating a frame-shift mutation. The CRISPOR analysis projected that the most likely mutation type is a 14bp deletion, based on sequence microhomology flanking the DNA target site.

I next compared the results of the CRISPOR analysis with the results from the MiSeq and CrispRVariants pipeline. The output of CrispRVariants provides the number of different reads generated by MiSeq and quantifies the occurrence of each read type. I manually evaluated the occurrence and types of reads occurring at *Top1* exon 15. I considered the position between nucleotide -3 and -4 at the sgRNA target as position "0", with the coordinates of the insertions or deletions of individual nucleotides 5' ("minus") or 3' ("plus") to this location (Figure 17a). In all three replicates, the three most commonly occurring mutations at *Top1* exon 15 were; -1:1bp insertion, -1:2bp deletion, and +1:1bp deletion, suggesting that CRISPR-Cas9 mutagenesis is stereotypic at this locus. In all three of these mutation types, a frame-shift is predicted in *Top1*. Interestingly, in all three replicates, the mononucleotide insertion in the -1:1I mutation was an adenine (A) nucleotide.

I evaluated the occurrence of these top three mutations as a percentage of the total number of MiSeq reads aligning to *Top1* exon 15. For each replicate, these three mutations collectively contribute 20.6%, 11.2% and 18.9% (mean 16.9%) of the total number of reads (Figure 17b). I also evaluated the occurrence of these three mutations as a percentage of total indel-containing reads. The percentage occurrence increases to 26.6%, 35.8% and 39.2% for each replicate (mean 33.9%, Figure 17c). Therefore, on average, over one third of the range of indel mutations can be pinpointed to just three precise frame-shift mutations occurring at the expected dsDNA target position.

Lastly, I assessed the occurrence of the *in silico* predicted mutation (14bp deletion) within these replicates. The most commonly occurring 14bp deletion was at nucleotide position -6, occurring with frequency of 3.6%, 4.3% and 6.0% (mean 4.6%) for each replicate (Figure 17d). This was quantified as a percentage occurrence of all reads containing indel mutations.

Overall, the high mutational efficiency of inducing mutations at *Top1* exon 15 by sgRNA1 CRISPR-Cas9 genome editing, enhanced by the stereotypic frame-shift mutations, led me to use the sgRNA1 plasmid to generate a stable transgenic mouse model.

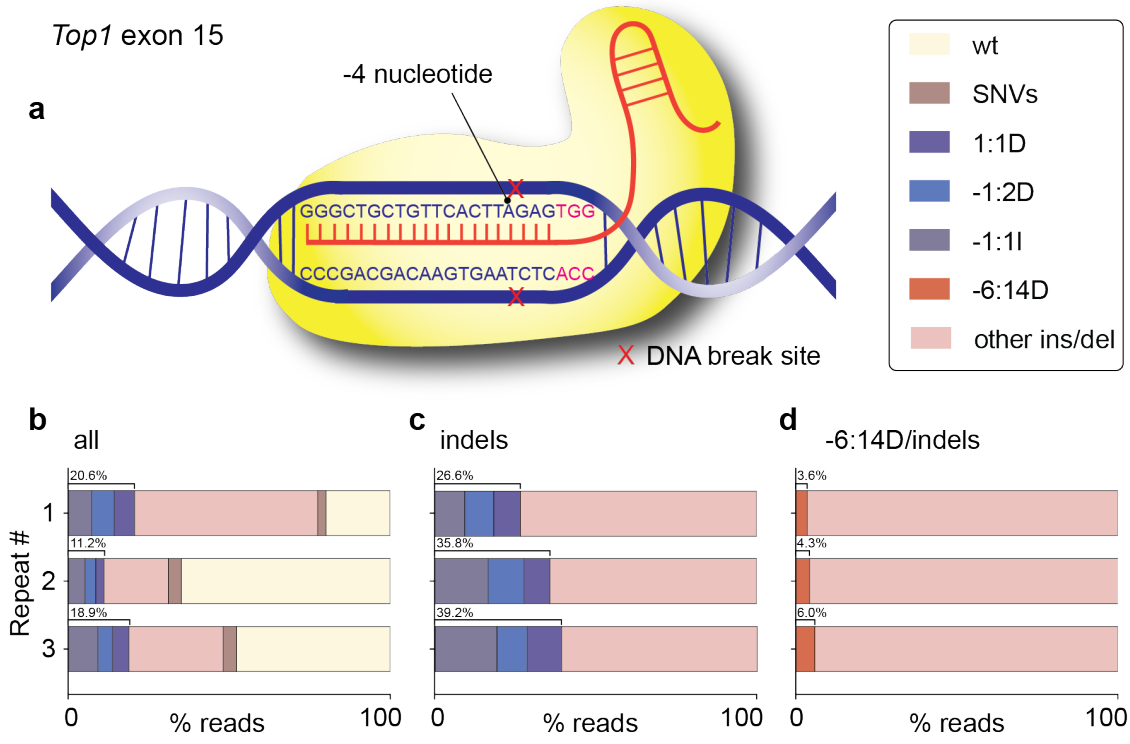

Figure 17. Dynamics of sgRNA1 mutations

(a) Schematic of CRISPR-Cas9 targeting at *Top1* exon 15, guided by sgRNA1, showing the sgRNA (blue) and PAM (pink) sequence. The Cas9 induces a DNA DSB at the -4 nucleotide position (red X), proximal to an A nucleotide. (b) The complete spectrum of read categories (wildtype, SNV, types of indel mutations). The dominant mutations (1:1D, -1:2D, -1:1I) are highlighted as a percentage of the total sequence reads. (c) The dominant mutations (1:1D, -1:2D, -1:1I) are shown as a percentage of the total reads containing indel mutations (d) The occurrence of a 14bp deletion at the -6 position as a percentage of indel mutations, a predicted highly dominant mutation by *in silico* tools; CRISPOR, crispor.tefor.net (Concordet and Haeussler, 2018, Haeussler et al., 2016).

4.2.6 Generating the sgRNA 1 knock-in mouse model

Given the highly mutagenic capacity of sgRNA1 with a prevalence of stereotypic frame-shift mutations, I next aimed to generate a transgenic mouse line that expresses sgRNA1. I was unable to use CRISPR-Cas9 genome editing to generate this knock-in, as any successful integrations of the transgene could have resulted in contemporaneous expression of residual Cas9 and the sgRNA transgene. The presence of both CRISPR-Cas9 components would target *Top1* for mutation, potentially resulting in non-viability of the successfully targeted pups. I therefore used a previously published knock-in approach called "TARGATT" (target-attP; Tasic et al., 2011) which utilises ϕC31 integrase-mediated recombination between attP and attB sequences. Recombination at the attP/B sites produces an attL/R sequence which is inhibitory to continued ϕC31 integrase action and prevents any further locus recombination. A mouse line containing three attP sequences inserted into the permissive H11 locus was previously generated ("H11-attPx3"; Tasic et al., 2011, Hippenmeyer et al., 2010). An attB-flanked transgene can be integrated into this transgenic H11 locus by microinjection of ϕC31 integrase and the targeting vector into H11-attPx3 embryos. The authors also published another mouse line where three attP sequences were integrated into *Rosa26* (*R26*-attPx3; Tasic et al., 2011). However for my sgRNA1 transgene targeting experiments, I decided to target the H11-attPx3 locus, as this provides future flexibility for transgenic mice to be mated with the *R26*-Cas9 mouse line (Platt et al., 2014).

I edited the *Top1* sgRNA1 pLethal plasmid (U6-sgRNA1:Cbh-mCherry) to contain attB sequences flanking the transgene (Figure 18a). The second, empty U6-sgRNA scaffold was retained, to ensure consistency within the structure of the construct if a future double sgRNA transgenic mouse line was made. The Cbh-mCherry cassette was kept for transgenic fluorescent reporter read-out.

To generate the stable transgenic mouse line, zygote stage hemizygous *H11*-attPx3 (Tasic et al., 2011) embryos were microinjected by the GeMS STP, with the knock-in

components; the attB-flanked targeting vector, and ϕC31 integrase mRNA, facilitating integrase-mediated recombination in the pronuclear DNA (Figure 18b).

All microinjected embryos that successfully cleaved to the two-cell stage were surgically transferred into the uteri of pseudopregnant females (n=5) by GeMS. Three of five females littered, giving a total of 25 pups born. I assessed the pups for presence of the transgene by *in vivo* imaging at 3 days post birth, which detected the expression of the mCherry reporter. Of 25 pups born, 1 male was positive, giving an knock-in efficiency rate of 4%. The one positive male was not confirmed to have a targeted integration at the *H11* locus.

The male founder was set up in matings with wildtype C57BL/6J females and offspring were assessed by *in vivo* imaging to identify the mCherry positive pups. In the first two litters produced, five pups were born that were mCherry positive (n=5/10, 50%). Therefore the founder successfully germline transmitted the transgene to the F1 generation (two F1 pups shown in Figure 18c).

To confirm mCherry expression at the RNA level, I performed a qPCR using a TaqMan probe. mCherry expression was assessed in four different tissues (brain, liver, kidney, spleen) dissected from an F1 generation sgRNA mouse, normalised to wildtype liver *Gapdh* expression. Results showed the mCherry was expressed in all four tissues, and not in the wildtype control (Figure 18d). The stable transgenic mouse line was annotated as "H11:U6-Top1sgRNA1:Cbh-mCherry", and is hereon referred to as "sgRNA".

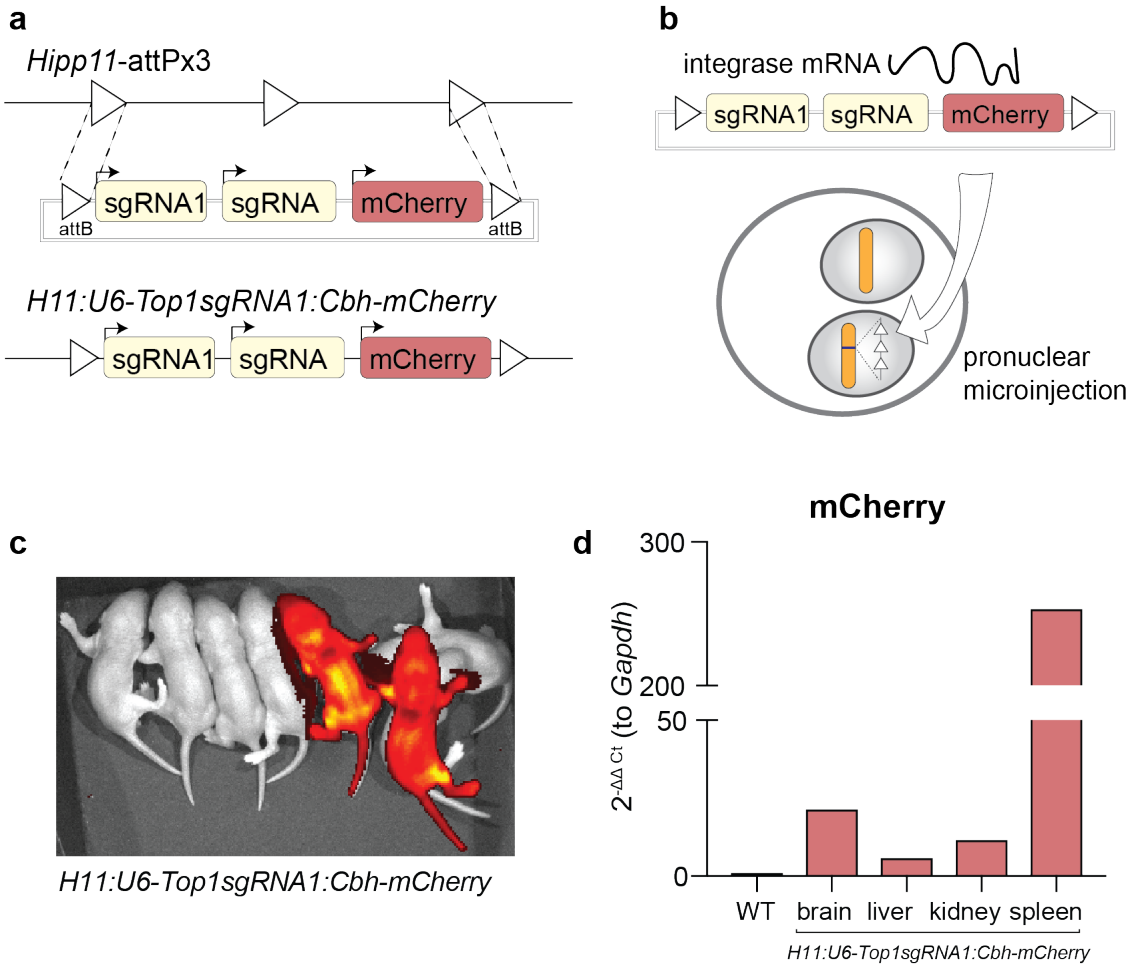

Figure 18. Generating an sgRNA expressing mouse line by TARGATT knock-in

(a) Schematic of the targeting strategy to generate a U6-sgRNA1:Cbh-mCherry knock-in at the permissive Hipp11 locus on autosomal chromosome 11 (H11). The pLethal plasmid described previously, containing a U6 promoter followed by sgRNA 1 (targeting Top1 exon 15) with a second tandem empty U6 promoter and sgRNA scaffold, followed by a Cbh promoter and mCherry reporter, was modified to contain attB sequences flanking the knock-in region of interest. (b) The attB-pLethal plasmid was co-injected with ϕC31 integrase into TARGATT H11-attPx3 heterozygote embryos. Successfully cleaved embryos were surgically transferred into pseudopregnant females to litter (c). A single male founder was generated and germline transmitted to the F1 generated, assessed by *in vivo* imaging. (d) The H11:U6-sgRNA:Cbh-mCherry ("sgRNA") mouse expressed mCherry in multiple tissues, quantified by qPCR, normalised to wildtype *Gapdh* expression. The embryo microinjections and surgical transfers were performed by the GeMS STP.

Once I determined that the sgRNA mouse line was stably transmitting the transgene and expressing the mCherry reporter, I assessed efficiency of sgRNA expression by mating with the *R26*-Cas9 mouse line, to model the bi-component CRISPR-Cas9 system *in vivo*. The *in vivo* bi-component system will be discussed in Chapter 5.

4.3 Discussion

In this chapter I derived a Cas9-expressing ESC line to routinely and easily evaluate the efficiency of sgRNAs *in vitro*, when the Cas9 and sgRNA are genetically segregated thereby modelling the bi-component system *in vitro*. I confirmed that the ESCs expressed Cas9 and eGFP, providing a useful fluorescent reporter system as a proxy for Cas9 expression. Three different sgRNAs targeting *Top1* were cloned into the pLethal plasmid vector, also expressing a mCherry reporter. This allowed me to perform a 'traffic light' system to sort and evaluate double positive (sgRNA+Cas9+) versus single positive (Cas9+ only) cell populations. Using ESCs provided an easily transfected *in vitro* model, with high numbers of cells in each sorted population for mutation and expression analyses.

I decided on a time point of 48 hours post-transfection for *Top1* sgRNA FACS, based on transfections with a non-lethal sgRNA, targeting *Sry*. Although the number of mCherry+ ESCs increased over time in the *Sry* transfection, the *Top1* sgRNAs may have a lethal phenotype in ESCs. Therefore, I chose 48 hours as a good compromise between high levels of transfected cells, and a reasonable time post-transfection to attempt to minimise ESC death. ESCs are usually passaged every 2-3 days, so 48 hours post-transfection also appeared a reasonable time according to standard ESC maintenance. When performing FACS, only the DAPI negative, alive ESCs, were collected for both mCherry+eGFP+ and eGFP+ only populations. The caveat of this approach is that any non-viable ESCs may be lost from the analysis. These ESCs may have large indel mutations that have resulted in an efficient loss-of-function phenotype and ESC lethality. This may be particularly relevant in the double sgRNA experiments, where sgRNA1 efficiency was lower post-transfection with double sgRNAs than with a single sgRNA1. Double sgRNAs may result in more complex mutations, or a "double hit" of indel mutations at two exons, resulting in loss-of-function of *Top1* more quickly. This caveat could be overcome by also collecting the DAPI positive population of ESCs. However since the reporter expression cannot be detected in this population, the cohort may have contributions of mCherry+eGFP+ ESCs and eGFP+ only cells. This would result in a highly inaccurate mutation efficiency calculation.

The mutation efficiency of sgRNA1 was higher than sgRNA2 and 3. In-depth analyses of the mutation dynamics of sgRNA1 showed a high likelihood of generating frame-shift mutations, increasing the probability of generating loss-of-function of *Top1*. Using ESCs to perform a low-throughput analyses to test multiple sgRNAs provided a simple and tractable method to assess mutation efficiency and make like-wise comparisons of each sgRNA, irrespective of transfection efficiency. One caveat to using ESCs to assess mutation efficiency is the number and variability of mutations present when transfecting large numbers of cells. In order to circumvent the manual process of individually aligning collapsed MiSeq reads to evaluate mutations at the target locus, I utilised the CrispRVariants R package (Lindsay et al., 2016). This allowed me to take into account every read aligning to the target site and accurately evaluate all of the variability in mutations in the target region. This also has the advantage of bioinformatically removing any non-specific or primer-dimer amplicons from the analysis, making analyses more streamlined.

I chose to evaluate the spectrum of mutations at a range of 20 nucleotides 5' and 3' from the 20bp sgRNA target. Any deletions up to 60bp long will be captured by this approach. Any mutations longer than 60bp are lost from the mutation efficiency quantification. To quality control check for large deletions, the data can be manually checked for large deletions greater than 60bp, and parameters of mutation efficiency quantification adjusted, if necessary. The maximum sequence length of the MiSeq amplicon is 500bp. Any mutations generating a product larger than 500bp would never be captured by PCR amplification, as the primer binding sites would be lost. This could be circumvented by using PCR primers that are complementary to sequence significantly further from the expected DSB site, and sequencing products smaller than 500bp.

Furthermore, there is a rising concern about large deletions and complex mutations arising using CRISPR-Cas9. Studies published in 2017 suggested that single sgRNAs induced deletions up to 600bp in mouse zygotes (Shin et al., 2017) and up to 1500bp in cancer cell lines (Gasperini et al., 2017). Kosicki *et al* showed that a single sgRNA targeting the X-linked gene *PigA* (Phosphatidylinositol glycan anchor biosynthesis class

A) in XY ESCs generated deletions up to 9.5kb (Kosicki et al., 2018, Allen et al., 2018). The potential for huge complexity in large deletions, inversions and rearrangements is not caught in this single sgRNA analysis.

To compare with the results of MiSeq/CrispRVariants, published *in silico* tools also give an indication of likely frame-shift mutational outcomes at the target site. The results shown by MiSeq/CrispRVariants at *Top1* exon 15 did not fully recapitulate the expected most commonly occurring mutation from these *in silico* tools. Single nucleotide variants were not considered to be 'true' mutations if they are not contained within the sgRNA target sequence. These may be naturally occurring single nucleotide variants. For example, a *R26*-Cas9 male was mated to MF1 females to collect embryos for ESC derivations. Given the different mouse genetic backgrounds, the ESCs could be variant rich. However, the occurrence of SNV-containing reads can be determined by this analysis and taken into account if required. In this analysis, I chose to keep SNVs as a separate category to indel mutations. In terms of variants, there may also be some PCR error or inefficiency, particularly towards the 5' and 3' edges of the amplicon during PCR amplification or sequencing. This can be circumvented by reducing the area flanking the sgRNA nucleotide sequence to investigate mutated reads, however this may lose large deletions in the region. The PCR amplicon generated for MiSeq sequencing may have some preferential bias, meaning that some sequences are amplified more readily than others. This would lead to an over-representation of some read types in the analysis. In order to try and reduce this amplification-bias, I kept the number of PCR cycles low, in order to amplify the region sufficiently for sequencing, but without over-amplification of few read types.

Importantly, with all of these described caveats, the eGFP+ only population was also included in the pipeline. This population undergoes the same PCR amplification, MiSeq sequencing and analysis by CrispRVariants and/or CRISPOR. Therefore this population highlights the basal variability around this locus, and after the analysis pipeline. In all of these analyses, I decided not to normalise to this population, to give a true indication of the variability at each *Top1* locus. As earlier described, there was a low non-Cas9-induced rate of mutations at the locus in these eGFP+ only ESCs. As well as the endogenous

variability at *Top1,* it is also possible that the pseudo-mutation rate is attributable to PCR error, during the amplification and library preparation steps. This PCR error would also occur in the mCherry+eGFP+ ESC populations in each experiment. Furthermore, the two populations are sorted by FACS on fluorescence reporter expression. It is possible that there was some contamination between population types during the collection, PCR amplification and library preparation steps. Contamination between samples would falsely increase the mutation efficiency in eGFP+ populations, and decrease the efficiency in mCherry+eGFP+ populations.

The data presented in this chapter showed that only when both CRISPR-Cas9 components are present, in the double positive ESCs, a high efficiency of mutations at the target site was introduced. I tested for TOP1 expression in transfected and non-transfected ESCs, at both the N-terminus and C-terminus of TOP1. This was because sgRNA1 and 2 target *Top1* at exons in the middle of the reading frame. Therefore, mutations in exon 15 or 16 may have generated a truncated protein which may have some function. Targeting by sgRNA3 at exon 2 near the expected start codon (ATG) may have induced a mutation resulting in alternative start codons being used. This means that a functional TOP1 protein may still be generated. In the single positive, Cas9+ only, populations TOP1 was expressed in all three samples, when using antibodies to both the N-terminus and C-terminus. Post-transfection with all three sgRNAs, in the Cas9+sgRNA+ ESCs, there was a loss of TOP1 protein, compared with the Cas9+ population.

Overall, this data highlights that a bi-component system efficiently drives mutations at *Top1* in a target-specific manner. The mutations induced at the target exon sites occurred after transient transfection with a plasmid *in vitro*. Therefore I speculate that the mutation efficiency may be greater in an *in vivo* environment due to constitutive expression of the sgRNA construct (when crossed with a constitutively expressing Cas9 mouse line). In the *in vitro* experimental strategy, three sgRNAs were tested that target the *Top1* locus. However this strategy is not limited to investigating mutational efficiency at this gene. This pipeline could be expanded to testing many sgRNAs targeting multiple genes, or the same gene in multiple loci. Furthermore, it is not limited to a single sgRNA but also multiple sgRNAs being expressed from a single plasmid, allowing for increased

flexibility in evaluating different knock-out designs. I hypothesized that a highly efficient sgRNA at the *in vitro* level, would translate to an *in vivo* approach and therefore generated the sgRNA expressing mouse line to test this hypothesis. I also aimed to investigate whether the most dominant mutational outcome at the *in vitro* level, would also be recapitulated *in vivo*.

Initially in the low-throughput sgRNA screen, I investigated efficiency of single sgRNAs and evaluated the mutations occurring at each target locus. This was a simple method of analysing the efficiency, spectrum and dynamics of the mutations for each sgRNA target. However given that there is also a high rate of mutagenesis when two sgRNAs are expressed in tandem, it may also be possible to generate knock-in mouse models expressing multiple sgRNAs. This may have a greater efficiency of generating a null mutation at the sgRNA target gene. Hypothetically, expression of the Cas9 and sgRNA in the early embryo could induce a mutation at the target locus that is a silent mutation. The presence of the silent mutation in the sgRNA target will prevent future DNA DSBs but not inhibit the protein function. Generating a mouse line that expresses multiple sgRNAs means multiple exons are targeted to induce frame-shift mutations, resulting in a greater likelihood of producing target gene loss-of-function.

A single sgRNA was shown to induce a high rate of mutations at the target locus in this data. Furthermore, single sgRNAs have been shown previously to have catastrophic deletion and rearrangements events in target genes (Kosicki et al., 2018). Using a double sgRNA system could further increase the possibility of a range of mutations or complex mutations in the intervening sequence between the two sgRNAs (Kraft et al., 2015, Boroviak et al., 2016, Boroviak et al., 2017). Complex mutations introduced by two/multiple sgRNAs are more challenging to accurately assess. This was shown in this results chapter when transfecting ESCs using the double sgRNA plasmids. Cas9-induced indels didn't always occur in the same DNA molecule. This data suggested that the two sgRNA/Cas9 complexes were not always cutting simultaneously at the same DNA molecule. There were also PCR products generated from large deletions of the intervening sequence in the double sgRNA transfections. These very large *Top1* deletions may contribute to an increased rate of ESC death. These non-viable, DAPI positive ESCs

would not have been sorted during FACS and therefore lost from the analysis. This may explain the underrepresentation of 'large deletion' clones found during the Sanger sequencing analysis. The 44kb distance between sgRNA1 and 3 could not be amplified by PCR and therefore the types of mutations occurring at each of the two sgRNAs in a single DNA molecule could not be evaluated.

Given the simplicity of screening for single sgRNA mutations, with *in silico* and bioinformatically predicted frame-shift mutations, the first priority was to generate a single sgRNA expressing mouse line. The sgRNA1 20 nucleotide *Top1* target site was sequenced conserved between mouse, cow, chicken and rat, and others. The bi-component *in vitro* system in mouse ESCs could be utilised for testing sgRNA mutation efficiency targeting genes of non-mouse species, provided they are sequence conserved.

Chapter 5. Results 3: *In vivo* bi-component CRISPR-Cas9 induced mutations at *Top1*

5.1 Introduction

The *in vitro* system described in Chapter 4 showed that expression of Cas9 and a *Top1* sgRNA in ESCs induced mutations at *Top1*. To become applicable to other species, for example agricultural animals, the bi-component system must be functional *in vivo*. There are two previously published studies of *in vivo* bi-component systems, one in the silkworm (Zhang et al., 2018), and one in the mouse (Yosef et al., 2019). The silkworm bi-component system used heterogametic females to induce female-specific lethality (Zhang et al., 2018). However, in the mouse, the males are the heterogametic sex. The Yosef *et al* (2019) mouse study used a Y-linked sgRNA transgene crossed with an autosome-linked Cas9. Both transgenes were inherited uniquely by males, and generated mutations in essential genes *Atp5b*, *Casp8* and *Cdc20* to induce male-specific lethality. The limitation of this study was the incomplete sex skew: 9 of 113 (8%) newborns were male. Although this skew was a statistically-significant deviation from the usual 1:1 ratio of males:females, it did not show completely all-female litters.

In this chapter I investigated the penetrance of lethality when the *Top1* sgRNA and Cas9 transgenes are co-inherited. I firstly examined whether mutations were detectable at the *Top1* locus in pre-implantation embryos and whether the mutations recapitulated the mutation spectrum seen *in vitro.* Secondly, I investigated whether co-inheritance of the transgenes resulted in embryonic lethality, and whether loss of Cas9/sgRNA embryos impacted the mean litter size.

5.2 Results

5.2.1 Assessing mutations at the *Top1* locus in pre-implantation embryos in the bi-component CRISPR-Cas9 system

In Chapter 4 I generated an sgRNA-expressing mouse line, *H11:*U6-*Top1*sgRNA1:Cbh-mCherry, hereon referred to as "sgRNA", targeting *Top1* exon 15. I hypothesised that inheritance of both the sgRNA and a Cas9 transgene *in vivo* would result in mutations in *Top1*, inducing embryonic lethality, whereas inheritance of either transgene in isolation would have no effect.

A traffic light system for detecting the sgRNA (mCherry) and Cas9 (eGFP) transgenes was utilised *in vitro* to sort the ESCs by genotype. Therefore, I first assessed whether the reporters were expressed in embryos. To pinpoint the time of transgene expression, I collected pre-implantation embryos at three stages: E1.5 (2-cell, n=9), E2.5 (8-cell, n=22) and E3.5 (early blastocyst, n=51) from timed mating between hemizygous sgRNA fathers (sgRNA/+), and homozygous *R26*-Cas9 mothers (Cas9, Platt et al., 2014; Figure 19a,b). There were two possible embryo genotypes; Cas9/sgRNA or Cas9/+. The Cas9/sgRNA expressed eGFP and mCherry, whilst Cas9/+ embryos only expressed eGFP. eGFP expression was detectable at E1.5, E2.5 and E3.5 in every embryo analysed. mCherry expression was absent from E1.5 embryos, but expressed at E2.5 and E3.5 (Figure 19c). Therefore, I confirmed that pre-implantation embryo genotypes could be assessed by reporter expression at E2.5 and E3.5. The early expression of eGFP but not mCherry in the E1.5 embryos was likely due to maternally deposited mRNA in the oocyte from the *R26*-Cas9 mother.

After sorting the embryos by reporter expression, the E2.5 and E3.5 embryos were lysed, PCR amplified for *Top1* exon 15 and sequenced by MiSeq. E1.5 embryos were not sequenced because they could not be genotyped by reporter expression and therefore any mutations at *Top1* could not be ascribed to genotype. For each embryo, the percentage mutation efficiency was calculated as the total number of reads containing an insertion or deletion mutation, divided by the total number of reads. Single nucleotide variant-containing reads were not counted as mutations. E2.5 Cas9/sgRNA embryos (n=14) had

a mean mutation efficiency at *Top1* exon 15 of 54.2%, whilst Cas9/+ embryos (n=8) had a mean mutation efficiency of 5.7%. The difference in mutation efficiency between Cas9/sgRNA and Cas9/+ embryos was statistically-significant ($p=3x10^{-3}$, Mann-Whitney test; Figure 19d). E3.5 Cas9/sgRNA embryos (n=16), had a mean mutation efficiency of 60.5% while in the Cas9/+ embryos (n=35) it was 7.4%. The difference in mutation efficiency between the Cas9/sgRNA and Cas9/+ embryos at E3.5 was statistically-significant ($p=1x10^{-10}$, Mann-Whitney test; Figure 19d). Similarly to in Chapter 4, there was a non-Cas9-induced rate of mutation in the Cas9/+ embryos, despite the supposed lack of sgRNA transgene. This mutation rate was most likely produced by PCR error during amplification and library preparation. Furthermore, the error may have also occurred by incorrect sorting of Cas9/sgRNA and Cas9/+ embryos, resulting in contamination of the 'incorrect' genotype contributing to the mean mutation efficiency. In the Cas9/sgRNA population, the presence of Cas9/+ embryos would falsely reduce the mean mutation efficiency, whilst in the Cas9/+ population, presence of Cas9/sgRNA embryos may false increase the mutation efficiency. Explanations for the mutation efficiency seen in the Cas9/+ population is described further in 5.3.

The mean mutation efficiency of Cas9/sgRNA embryos increased by 6.3% from E2.5 to E3.5, however this difference was not statistically-significant ($p=5x10^{-1}$, Mann-Whitney test). Overall, these results showed that co-inheritance of the Cas9 and sgRNA transgenes in pre-implantation embryos, generates CRISPR-Cas9 induced mutations at *Top1*.

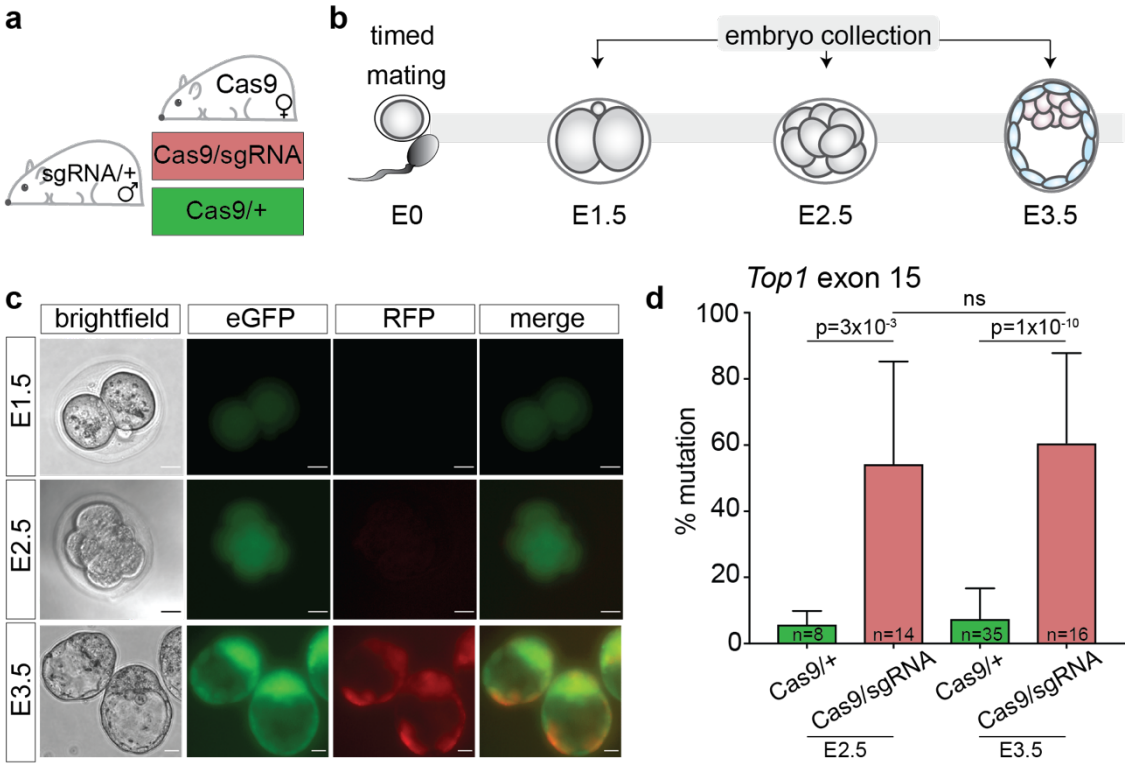

Figure 19. Evaluation of mutations at the *Top1* locus in pre-implantation embryos

(a) sgRNA/+ hemizygous male and *R26*-Cas9 homozygous female mice timed matings. (b) Embryos were collected at three pre-implantation stages; E1.5 (2-cell), E2.5 (8-cell) and E3.5 (blastocyst). (c) Embryos were imaged by fluorescence microscopy for mCherry and eGFP. Scale bars=20µm. (d) Genotype-sorted embryos were lysed, MiSeq sequenced and reads analysed by CrispRVariants. Statistical analysis was performed using Mann-Whitney tests. The MiSeq library preparations were performed by the GeMS STP. The MiSeq sequencing was performed by the ASF. The CrispRVariants R pipeline was performed by Jasmin Zohren (Turner lab).

5.2.2 Evaluating the spectrum of mutations

Co-inheritance of the Cas9 and sgRNA transgenes induced mutations at *Top1,* similarly to mCherry+eGFP+ ESCs *in vitro.* The most commonly occurring mutation in the *in vitro* screen was a 1bp insertion at the -1 position (-1:1I); which was consistently an A nucleotide. The A mononucleotide insertion was replicated in all three repeats of the *in vitro* experiments. Therefore, I assessed the occurrence of the -1:1I mutation in E2.5 and E3.5 Cas9/sgRNA embryos (E2.5, n=14; E3.5, n=16).

MiSeq sequencing of E3.5 embryos produced 61580 total mapped reads, on average, whilst E2.5 embryos produced an average of 42860 total mapped reads. Reads containing

the -1:1I mutation were present in 100% (n=30/30) of analysed embryos. In all 30 embryos (100%) the inserted mononucleotide in the -1:1I mutation, was also an A nucleotide. This accurately recapitulated the *in vitro* analysis. Next, I quantified the number of reads containing a -1:1I mutation as a percentage of the total reads for each embryo (Figure 20a). The prevalence of this -1:1I read type was varied. In E2.5 embryo #17 only 0.005% of reads contained the -1:1I insertion (4/73536 total reads, Figure 20a). Conversely, in E3.5 embryo #13, 76.3% of reads contained the -1:1I mutation (169264/221869 total reads, Figure 20a).

Next, I evaluated the number of reads containing the -1:1I mutation as a percentage of the total number of reads containing any indel mutation (Figure 20b). This quantification gives an indication of how frequent the -1:1I mutation is within the indel-containing read population. In E2.5 embryo #17, the -1:1I only occurred in 0.008% of reads (4/47342, Figure 20b). Conversely, E3.5 embryo #13 had 94.9% of -1:1I reads within the population of the indel-containing reads (169264/178378, Figure 20b). In 43.3% embryos (n=13/30; asterisk, Figure 20b), the -1:1I mutation was the most commonly occurring indel mutation.

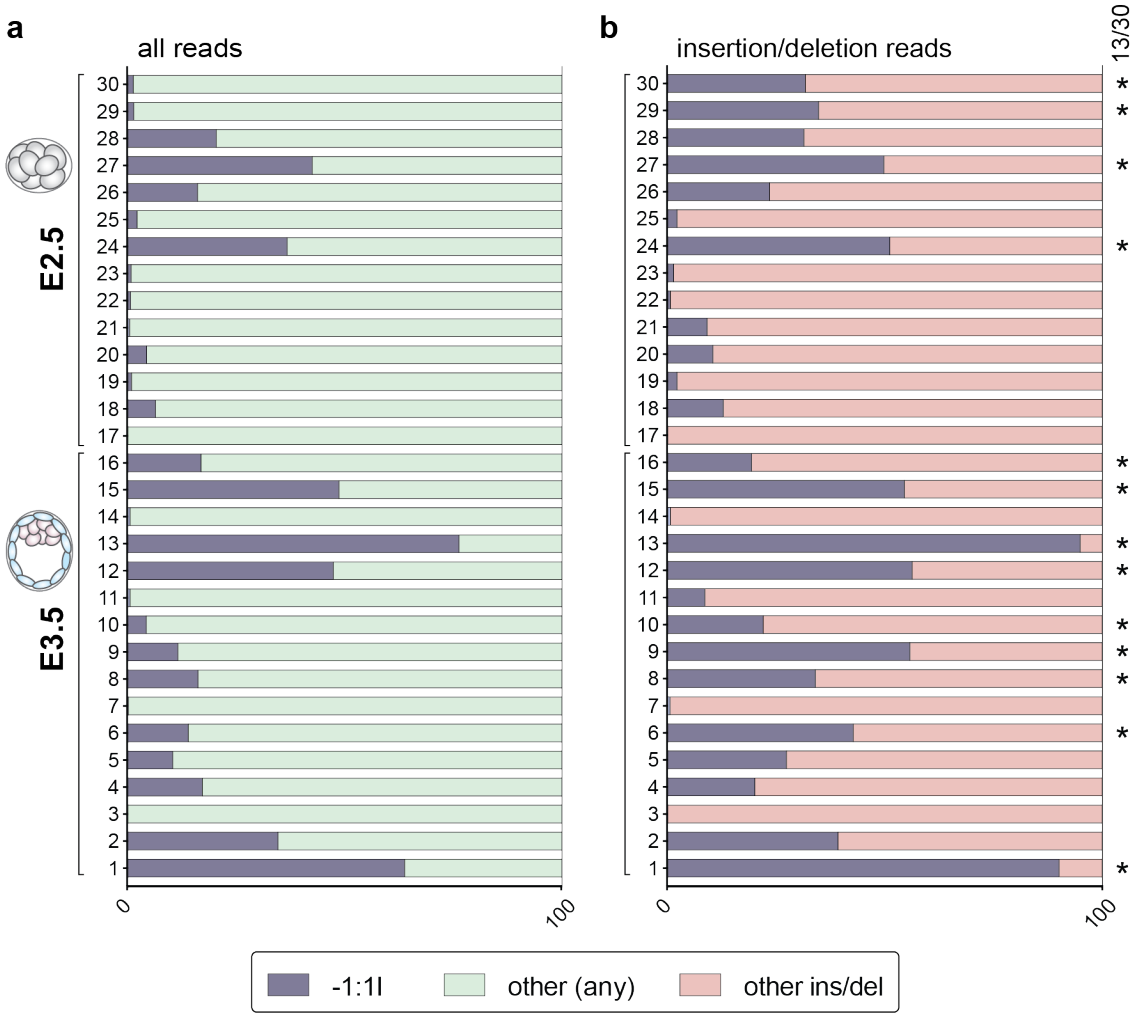

Figure 20. Occurrence of a -1:1bp insertion mutation in individual embryos

(a) The occurrence of a -1:1bp insertion read as a percentage of all MiSeq reads in each Cas9/sgRNA embryo at E2.5 (n=14) and E3.5 (n=16). (b) The occurrence of the -1:1bp insertion read as a percentage of all reads containing any insertion or deletion for the Cas9/sgRNA E2.5 and E3.5 embryos, as in (a). Asterisk indicates the individual embryo sample where the -1:1bp insertion read is the most commonly occurring/most dominant mutation (13/30).

I hypothesised that if the -1:1I mutation was the dominant mutational outcome, then the occurrence of the read would correlate with the total number of reads. I calculated the Pearson's correlation coefficient between the total number of reads, and the number of -1:1I reads per embryo. The correlation equalled 0.77, p=5.9e-07, a moderately high correlation. The correlation suggested that the -1:1I mutation was a likely outcome at *Top1* exon 15 induced by CRISPR-Cas9 mutagenesis. The raw read data per embryo is summarised in Table 1, highlighting the number of reads for the -1:1I mutation, the specific mononucleotide insertion, the total number of reads containing an insertion or deletion, and the total number of mapped reads generated by MiSeq sequencing.

Embryo #		-1:1bp insertion reads	Nucleotide	Total ins/del reads	Total reads
E3.5	1	53793	A	59724	84305
	2	32054	A	81603	92490
	3	53	A	76488	90443
	4	24578	A	121966	142352
	5	896	A	3269	8504
	6	880	A	2054	6251
	7	22	A	3019	6727
	8	1895	A	5560	11646
	9	538	A	965	4595
	10	321	A	1450	7217
	11	30	A	344	4168
	12	43975	A	78155	92822
	13	169264	A	178378	221869
	14	521	A	66209	79283
	15	24082	A	44176	49460
	16	14114	A	72612	83155
E2.5	17	4	A	47342	73536
	18	6954	A	53918	106973
	19	317	A	13498	31328
	20	1757	A	16757	38933
	21	291	A	3173	49498
	22	429	A	45978	52416
	23	362	A	24451	38317
	24	2081	A	4065	5652
	25	1485	A	63981	64091
	26	3063	A	12965	18859
	27	11924	A	23950	28009
	28	6205	A	19765	30210
	29	523	A	1505	35218
	30	383	A	1204	27000

Table 1. Occurrence of reads in Cas9/sgRNA embryos at E2.5 and E3.5

The raw number of reads for each Cas9/sgRNA embryo at E2.5 (n=14) and E3.5 (n=16). The table lists read occurrence for: -1:1bp insertion, the nucleotide inserted, total number of mutation reads, total number of all reads.

In conclusion, this data shows that a large percentage of Cas9/sgRNA embryos presented a stereotypic mononucleotide insertion mutation at the -1 position, as in the *in vitro* results. A mononucleotide insertion will induce a frame-shift mutation resulting in *Top1* reading-frame disruption. I therefore predicted that loss-of-function of *Top1* in pre-implantation embryos will induce non-viability, and loss of Cas9/sgRNA live births.

5.2.3 Cas9/sgRNA embryos are non-viable

To determine if CRISPR-Cas9 induced *Top1* mutations induced embryonic lethality, I assessed the frequency of genotypes born from the experimental matings. I set up matings between hemizygous sgRNA/+ fathers and homozygous *R26*-Cas9 mothers and genotyped their pups. Multiple breeding pairs were used to ensure there were no mating-pair specific effects. The experimental matings produced two possible offspring genotypes; Cas9/sgRNA and Cas9/+ (Figure 21ai). I hypothesised that the Cas9/sgRNA pups would not be born, due to *in utero Top1* loss-of-function induced lethality.

To assess the offspring genotypes, transgene-genotyping was performed on ear-biopsies from post-natal day 14 (PN14) pups. Results showed that there were no Cas9/sgRNA pups present from a total of 42 pups, from 7 litters (n=0/42, 0%; Figure 21b). This deviation from expected Mendelian ratios was found to be highly statistically-significant ($p=9x10^{-11}$, Chi-squared test).

Once I had confirmed that Cas9/sgRNA pups were not born, I compared the genotype ratios at PN14 to E2.5 and E3.5. At E2.5, 63% (n=14/22) of embryos were Cas9/sgRNA, whilst 37% (n=8/22) were Cas9/+. The genotype ratio was not a statistically-significant deviation from Mendelian frequency ($p=2x10^{-1}$, Chi-squared test; Figure 21b). At E3.5, 33% (n=16/48) embryos were Cas9/sgRNA, whilst 67% (n=32/48) were Cas9/+. The genotype ratio was a statistically-significant deviation from Mendelian frequency ($p=2x10^{-2}$, Chi-squared test, Figure 21b). This data showed that Cas9/sgRNA embryos were detected at E2.5 and E3.5 but not at PN14.

To confirm that Cas9/sgRNA embryos are non-viable, I switched the parental genotypes. Homozygous sgRNA females were set up in matings with hemizygous *R26*-Cas9/+ males (Figure 21aii). I genotyped pups from two litters at PN14 and results showed there were no sgRNA/Cas9 pups present (n=0/16, 0%; Figure 21c). This data confirmed that Cas9/sgRNA embryos are non-viable and this was a statistically-significant deviation from Mendelian frequency ($p=5x10^{-5}$, Chi-squared test).

To confirm that the sgRNA transgene in isolation was not causing the embryonic lethality, I set up heterozygous sgRNA/+ male mice in matings with wildtype C57BL/6J females (+/+, Figure 21d). In this mating there were two expected offspring genotypes; hemizygous sgRNA/+ or non-transgenic wildtype (+/+). Multiple breeding pairs were used to ensure there was no mating-pair specific effects . First, E3.5 embryos were collected and genotyped by reporter expression. Results showed that 56% (n=5/9) embryos were hemizygous sgRNA/+ embryos, whilst 44% (n=4/9) were wildtype (Figure 21e). At PN14, 33 pups were genotyped, 15 were sgRNA/+ (46%), whilst 18 were wildtype (54%, Figure 21e). This data confirmed that the sgRNA transgene in isolation is not inducing embryonic lethality as sgRNA/+ and +/+ offspring were present in approximately equal ratio.

5.2.4 Effect of Cas9/sgRNA lethality on litter size

The Cas9 and sgRNA transgenes induced embryonic lethality by *Top1* knock-out when co-inherited, whilst the Cas9/+ siblings were unaffected. I hypothesised that the loss of Cas9/sgRNA embryos would reduce the mean litter size by half relative to control matings. In these experiments, experimental matings were hemizygous sgRNA/+ males mated with homozygous *R26*-Cas9 females (Figure 21ai). The control matings were hemizygous sgRNA/+ males crossed with wildtype females (Figure 21d). Multiple breeding pairs were assessed to ensure litter size was not a mating-pair specific effect.

To determine the mean litter size for experimental and control matings, the number of pups per litter was counted at PN14. In the control mating, the mean litter size was 5.5 (n=33 pups from 6 litters, Figure 21f). In the experimental mating, the mean litter size

was 4.3 (n=26 pups from 6 litters, Figure 21f) Therefore, the experimental mating litter size was reduced by 22% relative to the controls. This reduction was not statistically-significant ($p=5x10^{-1}$, Mann-Whitney test). These results disproved my hypothesis and showed that lethality of Cas9/sgRNA embryos did not reduce the mean litter size by half.

5.2.5 Effect on sex ratios

Next, I investigated if there was an effect on offspring sex ratios in experimental matings compared to controls. I hypothesised that because both the Cas9 and sgRNA transgenes are autosome-linked, there would be no effect on offspring sex ratios. Hemizygous sgRNA/+ males were set up in mating with wildtype females as controls (Figure 21d), and homozygous *R26*-Cas9 females for experimental matings (Figure 21ai). I also set up homozygous sgRNA females with hemizygous *R26*-Cas9 males (Figure 21aii). The pups were phenotypically assessed at PN14.

In the control mating, 33 pups were phenotypically examined. Of the pups present, 22 were male (n=22/33, 67%), and 11 were female (n=11/33, 33%). The occurrence of each sex was not a statistically-significant deviation from the expected frequency ($p=6x10^{-2}$, Chi-squared test, Figure 21g). When hemizygous sgRNA/+ males were mated to *R26*-Cas9 homozygous females, 42 pups were examined. Of the pups present, 16 were male (n=16/42, 38%) and 26 were female (n=26/42, 62%, Figure 21g). The sex ratio in the experimental mating was a non-significant deviation from expected Mendelian frequency ($p=1x10^{-1}$, Chi-squared test). When hemizygous *R26*-Cas9/+ males were mated with sgRNA homozygous females, 7 of 16 pups were male (44%), whilst 9 of 16 were female (56%). This was a non-significant deviation from Mendelian frequency ($p=6x10^{-1}$, Chi-squared test, Figure 21g).

Overall, this data shows that the CRISPR-Cas9 bi-component system is able to induce embryonic lethality in embryos that have inherited both transgenes, but not when the sgRNA transgene is inherited in isolation. Although Mendelian inheritance frequency would predict the litter size to be halved, this was not the case.

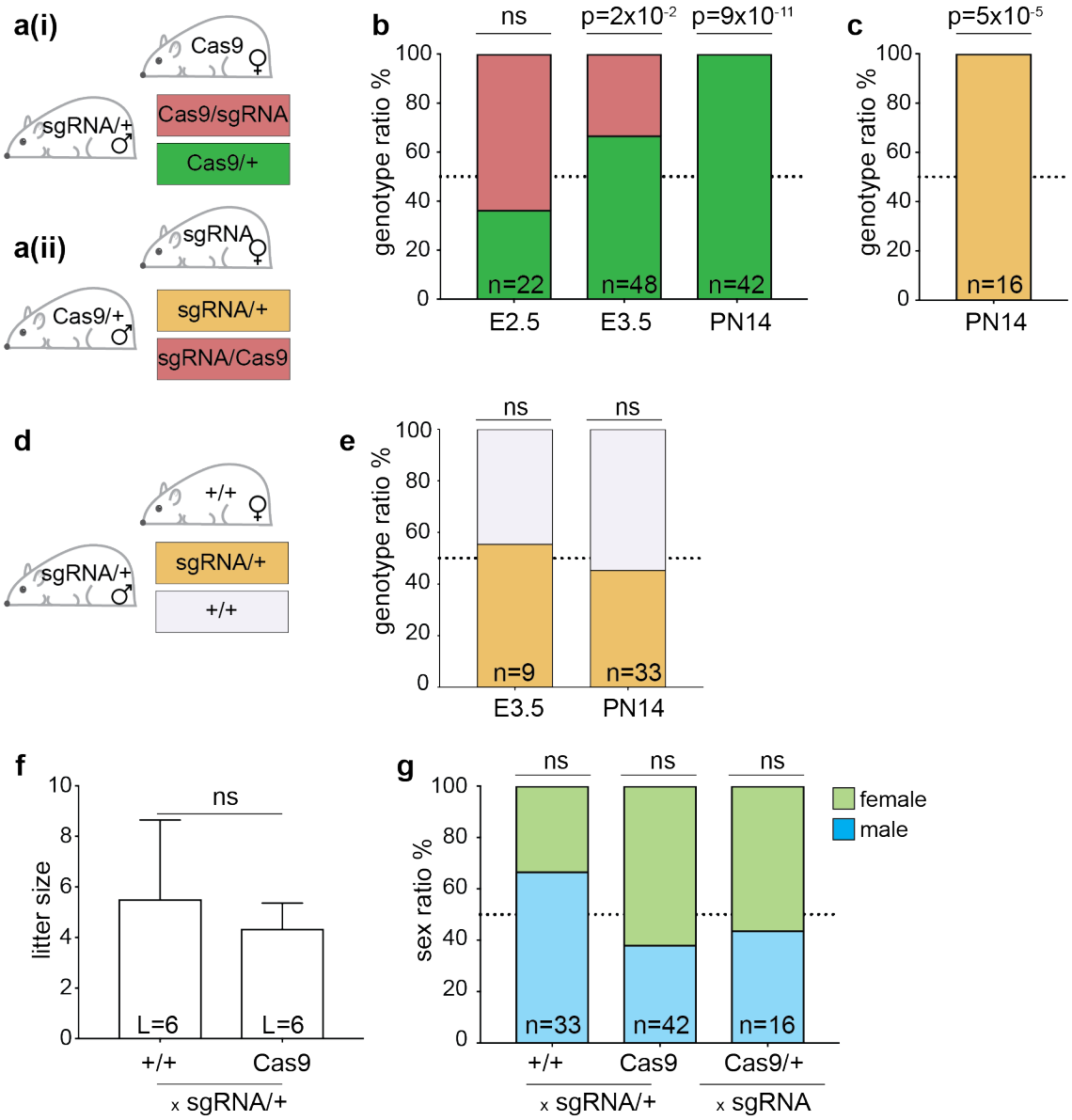

Figure 21. Developmental potential of transgenic sgRNA and Cas9 lines

(ai) Experimental mating strategy, crossing a heterozygous sgRNA/+ male with homozygous *R26*-Cas9 females. (aii) The reversed experimental mating strategy, crossing a homozygous sgRNA male with heterozygous Cas9/+ females. (b) The penetrance of each offspring genotype at E2.5 (8-cell), E3.5 (blastocyst) and post-natal day 14 (PN14) from mating (ai). (c) The penetrance of each offspring genotype at PN14 from mating (aii). (d) A control mating to assess sgRNA allele induced lethality; crossing a heterozygous sgRNA/+ male with C57BL/6J wildtype females. (e) Assessing genotype ratio from mating (d). (f) Mean litter size from control matings (d) and experimental mating (ai). Statistical analysis with Mann-Whitney test. (g) The sex ratio of offspring from three different mating strategies; sgRNA/+ by wildtype (c), sgRNA/+ by *R26*-Cas9 (ai) and sgRNA by *R26*-Cas9/+ (aii). Deviation from expected ratios measured using Chi-squared tests.

5.2.6 Post-implantation (E11.5) assessment of Cas9/sgRNA embryos

Previous studies have shown that *Top1* knock-out is embryonic lethal in the mouse at the 4-16 cell stage (Morham et al., 1996). Although Cas9/sgRNA embryos were not born, I was able to harvest Cas9/sgRNA blastocyst-stage embryos. Therefore, I determined whether the Cas9/sgRNA embryos could also be detected post-implantation. I therefore collected and genotyped embryos from hemizygous sgRNA/+ by homozygous *R26*-Cas9 matings at E11.5 (Figure 22a).

At E11.5, I collected 33 embryos from 4 matings. Of these, 23 were easily dissected from the uterus. Ten of the embryos (n=10/33, 30%) had been reabsorbed and could not be genotyped (Figure 22b). All 23 developing embryos (n=23/23, 100%) expressed eGFP, but did not express mCherry (representative image in Figure 22d). The 23 surviving embryos were genotyped for the Cas9 and sgRNA transgenes. Results confirmed that all of the embryos were Cas9/+ (Figure 22c). This data suggests that Cas9/sgRNA embryos have been lost by E11.5, and some of the reabsorbed embryos may have been Cas9/sgRNA.

In conclusion, the sgRNA transgene targets *Top1 in vivo* and commonly produces a single mononucleotide A insertion at the minus 1 position resulting in a frame-shift mutation. Co-inheritance of the Cas9 and sgRNA alleles induces embryonic lethality due to *Top1* loss-of-function, as highlighted by a complete loss of Cas9/sgRNA embryos by E11.5. To generate sex-specific lethality, I looked to generate a sex chromosome-linked Cas9 transgene, described in the next Chapter.

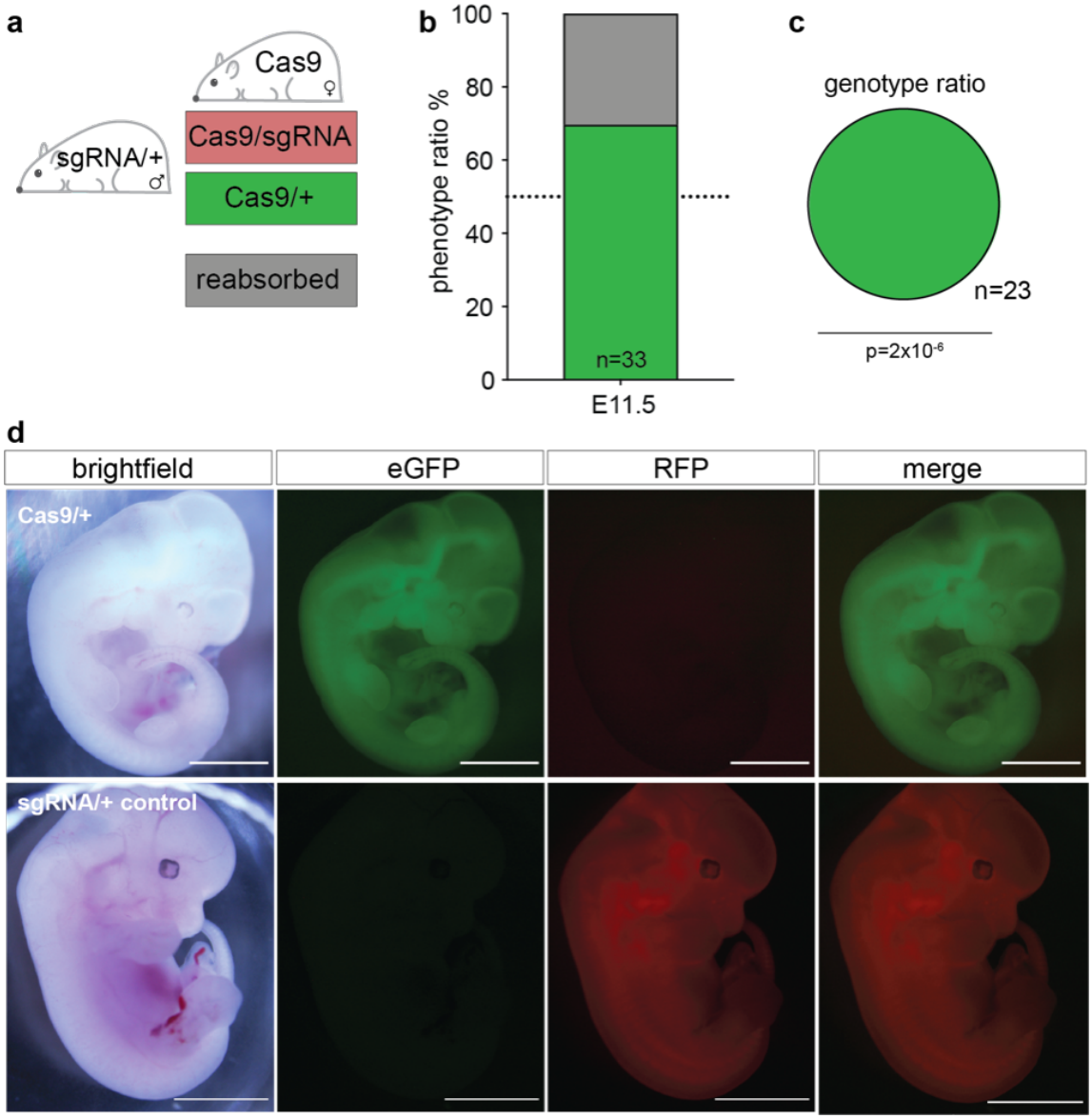

Figure 22. Assessment of E11.5 embryos

(a) Hemizygous sgRNA/+ males were crossed with homozygous *R26*-Cas9 females ("Cas9") and embryos collected at E11.5. (b) Assessment of phenotype by fluorescence microscopy by reporter expression, eGFP or RFP. (c) Genotyping for Cas9-eGFP or sgRNA-mCherry. Statistically-significant deviation from expected offspring genotypes (Chi-squared test, $p=2x10^{-6}$). (d) Upper panel: representative E11.5 embryo fluorescence microscopy and brightfield images. Lower panel: a sgRNA/+ heterozygous embryo was included as a positive control for mCherry fluorescence. Scale bar=1mm.

5.3 Discussion

The data presented in this chapter highlighted the capacity for the CRISPR-Cas9 bi-component system to induce mutations *in vivo*. The mutation rate was significantly higher in Cas9/sgRNA embryos compared to Cas9/+ at both E2.5 and E3.5. Moreover, the Cas9/sgRNA embryos were non-viable, and could not be recovered by E11.5.

To determine the different genotypes, embryos were phenotyped by the expression of eGFP and mCherry reporters. The eGFP was detectable by fluorescence microscopy as early as E1.5, the 2-cell stage. This is unsurprising because the *R26*-Cas9 allele is maternally-inherited. Maternal transcripts are deposited in the oocyte for early embryonic development post-fertilisation, prior to EGA (Clegg and Piko, 1983, De Leon et al., 1983, Wassarman and Kinloch, 1992). Cas9 transcripts are also deposited in the early embryo, when the mother encodes a Cas9 transgene (Cebrian-Serrano et al., 2017). Therefore, it is likely that eGFP expression would be detected in the 2-cell stage before the onset of EGA (Flach et al., 1982, Bernstein and Mukherjee, 1972). To confirm if eGFP is maternally loaded, oocytes and zygote stage embryos should be checked for eGFP expression. The eGFP reporter remained expressed at all assessed pre-implantation stages of development and at E11.5. Expression of mCherry was not visible at E1.5, but expressed at E2.5, suggesting transgene activation occurs at EGA. By the early blastocyst-stage, mCherry was clearly expressed, and the expression remained at E11.5 and after birth.

Once the embryos were sorted by genotype, I assessed the mutation efficiency at the *Top1* locus. *Top1* knock-out in the mouse induces embryonic lethality at the 4-16 cell stage (Kobayashi et al., 2011, Morham et al., 1996, Wright et al., 2015). In these previously published studies the knock-outs were generated by mating two heterozygous knock-out parents. In my experiments, Cas9/sgRNA embryos developed to E3.5. There are a number of possible explanations for this extended survival. The CRISPR-Cas9 system takes longer to become active and generate mutations at the *Top1* locus to induce non-viability, compared to inherited homozygous knock-out alleles. Although Cas9 mRNA was likely maternally-loaded, the sgRNA-transgene was not activated until post-EGA,

therefore formation of CRISPR-Cas9 induced mutations may be delayed. The embryo may have undergone the first cleavage divisions during the period of inactive CRISPR-Cas9, utilising *Top1* maternal mRNA (Clegg and Piko, 1983, De Leon et al., 1983, Wassarman and Kinloch, 1992). By E2.5, the CRISPR-Cas9 system was active, and the mean mutation efficiency was 54.2%, suggesting that approximately half of the *Top1* alleles contained a potentially loss-of-function mutation. By E3.5, the mean mutation efficiency increased to 60.5%, suggesting that the majority of alleles contained a potentially disruptive mutation.

The mutation rate for E2.5 and E3.5 embryos is calculated as the percentage of indel-containing MiSeq reads from the total number of reads. The mean mutation rate for both E2.5 and E3.5 suggests that the majority of reads contain a potentially loss-of-function mutation. The indel-containing reads contributing to the mutation efficiency calculation are not sorted into in-frame and frame-shift mutations. Therefore, it is possible that indel-mutations that are in-frame and would not induce loss-of-function. The 'loss-of-function mutation rate' could be calculated from the percentage of frame-shift indel-containing mutations, as a percentage of the total number of reads. Although Cas9/sgRNA embryos were present at E3.5, they were increasingly lost from the population at each stage. At E2.5, Cas9/sgRNA embryos made up 63% of the litters. By E3.5, this was reduced to 33%, and was 0% by E11.5. The loss of Cas9/sgRNA embryos over developmental time suggested there was some lethality at each stage. The Cas9/sgRNA embryos being lost at each stage may have had a greater mutation efficiency than the average mutation capture by MiSeq analysis, and were degraded before harvesting and therefore were not included in the analysis.

Once the mutations have been made, any remaining TOP1 protein previously translated has to be degraded from the developing embryo, before a complete loss-of-function has been generated. The predicted TOP1 half-life varies between different cell types. Kobayashi *et al* (2009) estimated TOP1 half-life is approximately 3.7 hours in limb bud cells (Kobayashi et al., 2009), whilst Desai *et al* (1997) estimated 10-16 hours in FM3A and ts85 mouse mammary carcinoma cells (Desai et al., 1997). All of these factors may contribute to the delay in non-viability compared to previous knock-out experiments.

Similarly to the *in vitro* analysis in Chapter 4, the Cas9/+ embryos show a small mutation rate at the *Top1* locus. The non-Cas9-induced pseudo-mutations seen are most likely due to PCR or sequencing errors during the MiSeq PCR amplification and sequencing pipeline. The Cas9/+ mutation efficiency is an internal control for basal variability at the *Top1* locus and therefore I did not normalise to this value. The E2.5 and E3.5 embryos were segregated into Cas9/sgRNA or Cas9/+ genotype by the expression of mCherry and eGFP reporters. Sorting at E2.5 by reporter expression was more challenging, because mCherry was lowly detected by fluorescence microscopy. Therefore, it is possible that there was some 'contamination' by incorrect embryo sorting (i.e. an mCherry+eGFP+ embryo may have been considered to be eGFP+ and vice versa). To overcome these genotyping challenges associated with early reporter expression, reporters could be driven by earlier pre-implantation gene promoters.

The small number of cells in E2.5 and E3.5 embryos meant that it was not possible to perform genotyping PCRs prior to MiSeq PCRs for more accurate segregation of genotypes. Despite this caveat, the difference in *Top1* mutation efficiency between Cas9/+ and Cas9/sgRNA embryos was highly statistically-significant, both at E2.5 and E3.5.

The Cas9/sgRNA embryos were non-viable from E11.5, whilst their sgRNA/+ littermates were viable. I predicted that the mean litter size of the experimental matings would be half of the control matings. However the experimental mating only showed a 22% reduction in mean litter size compared to the control. These results recapitulate the litter size results seen in Chapter 3, using the Cre-inducible DTA. The X-Cre male mating to DTA homozygous females induced a complete sex skew, generating all-male litters. Therefore, I predicted that the mean litter size would be halved. On the contrary, the mean litter size was reduced by a non-significant value of 32%, compared with the control. Curiously, the non-significant reductions seen in both the Cre/DTA and CRISPR-Cas9 approaches conflicts with the results in the Yosef *et al* study, the only other mouse bi-component study published thus far. Yosef *et al* reported a control mean litter size of 6.57 and an experimental mean litter size of 3.71 (Yosef et al., 2019). Although the

discrepancies in mean litter size can be partially due to different mouse genetic backgrounds, the non-viability effect appears to be stronger in the Yosef *et al* study.

I speculated that there may be multiple factors or events that are contributing to the reduction but not halving of the mean litter sizes seen in my experimental matings. In the control matings, e.g. sgRNA/+ by +/+, there could be some embryos that do not develop at a synchronous rate to siblings, and/or are unable to implant. These embryos may be normally non-viable. Furthermore, it is possible that in these control matings, there is an optimised number of embryos that have enough space to develop *in utero* successfully. In the experimental matings, some loss of the embryonic lethal genotype prior to implantation may allow embryos that usually cannot implant, to implant successfully. The reduction in Cas9/sgRNA embryos from E2.5 to E3.5 suggests that some are being lost between these two embryonic stages. The loss of E2.5 embryos may allow embryos that were not destined to implant, to do so. It is likely that both embryos types could 'attempt' to grow and implant but given that Cas9/sgRNA is embryonic lethal, only the single-positive embryos would survive. Therefore, the litter size is slightly, but not entirely, compensated. This partial compensation would result in a reduction, but not halved litter size of the experimental compared to the control.

The partial compensation by further embryo implantations may only be possible if the non-lethality effect is induced during pre-implantation, as in my experiments. In the Yosef *et al* study they targeted genes *Atp5b, Casp8* and *Cdc20. Atp5b* and *Casp8* knock-out embryos are embryonic-lethal at E9.5 and E10.5, respectively. However *Cdc20* knock-out is embryonic lethal prior to E3.5. If *Cdc20* was not disrupted, and loss-of-function mutations were induced only at *Atp5b* and/or *Casp8,* knock-out embryos would be non-viable post-implantation. Therefore it is unlikely any partial compensation pre-implantation mechanisms would be functioning to replace lost embryos.

The partially compensated litter size may be because the pregnant females induced diapause during gestation. Diapause is a process of embryonic 'suspension' at the blastocyst-stage, where the embryo is held in a period of non-development. Embryonic diapause in the mouse can occur from one day to several weeks (Weichert, 1940,

Weichert, 1942, Renfree and Fenelon, 2017, Pritchett-Corning et al., 2013). It is possible that mating of the CRISPR-Cas9 lines and resultant non-viability in some offspring may have induced an *in utero* stress. The massive embryonic lethality may be sufficient to induce the female into diapause. Holding the viable embryos in diapause may allow for some compensation of delayed embryos to also implant and develop. It is again possible that both delayed Cas9/sgRNA and viable genotype embryos may attempt to implant, but only viable embryo genotypes will survive. Therefore, there will be partial compensation but not complete compensation. However, diapause is usually induced in response to the presence of suckling pups, preventing the oestrogen surge and thereby causing the blastocysts to enter diapause (Renfree and Fenelon, 2017). In 5/6 (83%) litters measured for litter size quantification, the female did not have suckling pups during the pregnancy.

A further possible explanation to the partial compensation of litter size is the non-random fertilisation of gametes. A study in 2017 reported that mutations in 12 different genes resulted in a significantly lower or greater number of heterozygotic offspring than expected, and was entirely deficient in homozygotes. However, in each of the 12 genes studied, there was no reduction in litter size. Similar evidence was shown for mutations in genes affecting neural tube development, such as *Apob* (Nakouzi and Nadeau, 2014, Nadeau, 2017). Mice deficient for the genes were embryonic lethal, and genotype ratios were non-Mendelian. Upon addition of folic acid to the diet, the loss-of-function phenotype was rescued, and knock-out pups were born at Mendelian frequency. In each case, the mean litter size was similar (Nadeau, 2017). These studies suggested that there is a mechanism of non-random fertilisation of the gametes by transmission ratio distortion (TRD). Lyon described TRD as a departure from expected 1:1 offspring genotype ratio (Lyon, 2003). It is thought that TRD in females occurs by preferential entry of one allele into the polar body at meiosis, whilst in males, likely through sperm dysfunction. An example of TRD that affects sperm function is expression of the t-complex, described in 1.6.3.2 (Lyon, 2003). The litter size experiments in this thesis used a homozygous *R26*-Cas9 female, therefore all of the gametes would be equal. However the male was a hemizygous sgRNA/+ mouse. Therefore, it is possible that the + sperm carry a preferential fertility advantage over sgRNA-carrying sperm. This explanation for compensated litter sizes is unlikely, because in the control matings, both + and sgRNA-

carrying sperm were fertilised + oocytes to generate sgRNA/+ and +/+ offspring at approximately equal ratios. Therefore, the + sperm do not appear to carry a fertility advantage.

The generation of a litter that is embryonic lethal for the unrequired genotype but does not reduce the litter size by half is extremely advantageous for translation of the technology. Although in the dairy cow industry the number of offspring per litter is usually one, other agricultural species such as chickens and pigs usually have larger litters. A partial compensation of the litter size using this CRISPR-Cas9 bi-component technology would be extremely economically advantageous as a greater number of the required sex would be produced, compared to non-sex selected breeding. Furthermore, this would also be advantageous in the research laboratory setting, where a greater number of the required sex can be produced in a single litter. This could reduce the number of overall breeding pairs that would need to be set up in matings to generate the required offspring, thereby aligning with the Home Office 3Rs.

Chapter 6. Results 4: Generating sex chromosome-linked Cas9 knock-ins

6.1 Introduction

Generating a knock-in on the mouse X or Y chromosome ensures sex-specific inheritance of the transgene. If carried by the heterogametic male in a mouse mating, a transgene integrated on the X chromosome will be uniquely inherited by the daughters, and a Y-linked transgene, by the sons. When an X- or Y-linked Cas9 transgenic male is crossed with a homozygous sgRNA-expressing female, the inheritance of both CRISPR-Cas9 components is sex-specific. Co-inheritance of a Cas9 and sgRNA allele induces mutations at *Top1,* resulting in *Top1* loss-of-function and embryonic non-viability.

In this Chapter I aimed to generate an X- and Y-linked Cas9 transgene by *in vitro* targeting of ESCs. Successfully targeted ESCs could be used to generate stable transgenic mouse lines. In Chapter 4 and Chapter 5 I utilised mCherry and eGFP reporters for read-out of the sgRNA and Cas9 transgene expression, respectively. Therefore, I aimed to recapitulate this strategy by using a Cas9-eGFP transgene targeting the X or Y chromosome, where eGFP could be used as a read-out for Cas9 expression.

In Chapter 3 I used an X-linked Cre transgene (Tang et al., 2002) to induce DTA expression in female embryos. I was initially concerned that the paternally-inherited X-Cre allele would be transcriptionally silenced in female pre-implantation embryos due to imprinted X-chromosome inactivation in the mouse. However using a Cre reporter TdTomato line, I showed that the X-Cre is active by at least E2.5. Therefore, I aimed to recapitulate the targeting strategy of X-Cre to generate the X-Cas9 transgenic mouse line. I targeted the X-linked locus permissive loci, *Hprt,* as *Hprt* knock-out has no known detrimental phenotype in the mouse (Kuehn et al., 1987, Hooper et al., 1987, Koller et al., 1989, Jinnah et al., 1990).

Conversely to the X-linked *Hprt* gene, the Y chromosome does not contain any known permissive loci. Therefore, finding a locus to integrate the Cas9 transgene was

significantly more challenging. Many of the Y-genes have essential functions in endogenous cell mechanisms, for example DNA replication and repair (Lahn and Page, 1997, Bellott et al., 2014, Cortez et al., 2014, Bachtrog, 2014) or in sex determination and spermatogenesis, for example *Sry* (Gubbay et al., 1990, Koopman et al., 1991) and *Eif2s3y* (Yamauchi et al., 2014, Mazeyrat et al., 2001).

Although these Y-linked genes cannot be used as permissive loci to integrate transgenes into, some Y-gene promoters have been utilised to drive transgene expression. For example, studies have generated transgenic *Sry* reporter mouse lines, with *Sry*-promoter driven eGFP, such as the C57BL/6-Tg(Sry-GFP)92Ei Chr $Y^{AKR/J}$ /EiJ ("Sry-GFP") mouse line (Dewing et al., 2006, Albrecht and Eicher, 2001). In 2013, Wang *et al* generated knock-ins on the Y chromosome, utilising a method of genome editing called TALENs (Transcription activator-like effector nucleases; Wang et al., 2013a). They successfully targeted *Sry* for knock-out using two pairs of TALENs, specifically targeted to the high mobility group (HMG) DNA binding domain of the gene sequence. The XY offspring were sex-reversed (Wang et al., 2013a). They also used TALENs to generate an eGFP knock-in at the *Sry* locus in mouse ESCs. Targeted ESCs were injected into embryos and the injected embryos were left to develop *in utero*, before assessment at E12.0. Transgenic eGFP expression was visible in the gonad and brain (Wang et al., 2013a). Imaimatsu *et al* generated a flag-tag knock-in at the C-terminal end of the *Sry* gene using CRISPR-Cas9 and reported normal testis differentiation and spermatogenesis (Imaimatsu et al., 2018). The disadvantage of *Sry*-promoter driven transgenic lines is that *Sry* expression in the mouse is very tightly controlled from E10.5 to E12.5 during embryonic development (Koopman et al., 1990, Kashimada and Koopman, 2010) resulting in limited expression of the transgene.

Wang *et al* (2013) also generated a *Uty*-linked eGFP reporter line. *Uty* is ubiquitously expressed in the embryo, and *Uty* knock-out appears to have no detrimental effect on viability or fertility (Shpargel et al., 2012). *Uty* is expressed in ESCs, allowing for characterisation of reporter integration and expression *in vitro*. A *Uty*-eGFP ESC line was successfully generated and expressed eGFP. However the study did not comment on the effect on *Uty* expression after integration of eGFP (Wang et al., 2013a). This study

highlighted the possibility of generating expression constructs on the Y chromosome and *Uty* may be a useful target for integrating a Cas9-GFP transgene, given the ESC expression.

To generate single-sex litters, the components of the genetically segregated CRISPR-Cas9 system must be expressed during pre-implantation development. Pre-implantation expression maximises the length of developmental time the Cas9 and sgRNA transgenes are expressed to generate null mutations at the target locus, for example at *Top1*. Driving Cas9 expression by a male-determining gene promoter such as *Sry* means that the Cas9 expression window is short, and may be insufficient for Cas9 activity. Therefore a constitutive CAG promoter was utilised for both the X- and Y-Cas9 targeted lines.

6.2 Results

6.2.1 CRISPR-Cas9 HDR components for X-Cas9 HDR

The first aim of this Chapter was to generate a constitutively expressing Cas9-eGFP knock-in to the X chromosome *Hprt* locus, hereafter called "X-Cas9" (Figure 23a). I generated an X-Cas9 targeting vector (TV) encoding Cas9 and eGFP linked via a T2A sequence. The pX458 plasmid (Ran *et al,* 2013) was used as the dsDNA plasmid backbone. *Hprt* homology arms were amplified using gDNA extracted from C57BL/6J ESCs. The 5' homology arm (3.2kb; HA^L) was inserted into pX458 using unique restriction sites XbaI and KpnI. The 3' homology arm (1.1kb; HA^R) was inserted into the pX458 plasmid using unique restriction sites NotI and NarI. Both homology arms were confirmed by Sanger sequencing, as any mismatches reduce the efficiency of homologous recombination. Successful integration of the X-Cas9 TV by HDR deletes 173bp of *Hprt,* resulting in *Hprt* loss-of-function. A PGK promoter driving a neomycin resistance cassette, flanked by LoxP sites, was inserted into the X-Cas9 TV by Gibson Assembly cloning (Figure 23a). The neomycin gene allowed for selection of transgenic ESC clones.

Once I generated the X-Cas9 TV, I evaluated if eGFP was expressed from the TV *in vitro*. Wildtype ESCs were transiently transfected with either the X-Cas9 TV, the pX458 backbone plasmid, or were left non-transfected. Two days post-transfection, the ESCs were examined by fluorescence microscopy for eGFP expression. In the untransfected ESCs there were no cells expressing eGFP, however in the X-Cas9 TV-transfected ESCs, eGFP was expressed in some cells (Figure 23b). To confirm the X-Cas9 TV expressed Cas9, the essential component of the transgene, protein was extracted from bulk ESCs and analysed by western blot. Transfection of ESCs with pX458 or the X-Cas9 TV showed that Cas9 was expressed from the plasmids (Figure 23c). There was no Cas9 expression in the untransfected control ESCs. These results show that the X-Cas9 TV expresses Cas9 and eGFP and therefore could be used as the *Hprt* knock-in vector.

To maximise the efficiency of generating an X-Cas9 knock-in at *Hprt*, I generated a CRISPR-Cas9 sgRNA-expressing plasmid which induces a DNA DSB at *Hprt* exon 2 (Figure 23a). The sgRNA targeted the 173bp intervening DNA sequence between the two homology arms. The 20 nucleotide sgRNA oligonucleotide was cloned into pX458, using unique restriction site BbsI and was confirmed by Sanger sequencing. To confirm that the sgRNA induces DNA DSBs at *Hprt*, the sgRNA-pX458 plasmid was transiently transfected into wildtype ESCs. pX458-encoded eGFP expression was used to sort transfected ESCs from non-transfected ESCs by FACS. The populations of eGFP+ and eGFP- ESCs were lysed, the *Hprt* region amplified by PCR and sequenced by MiSeq. The MiSeq reads were analysed by the CrispRVariants pipeline to determine the mutation efficiency at the *Hprt* locus, which measured 66.0% (Figure 23d). I concluded that the sgRNA induces DNA DSBs at *Hprt*.

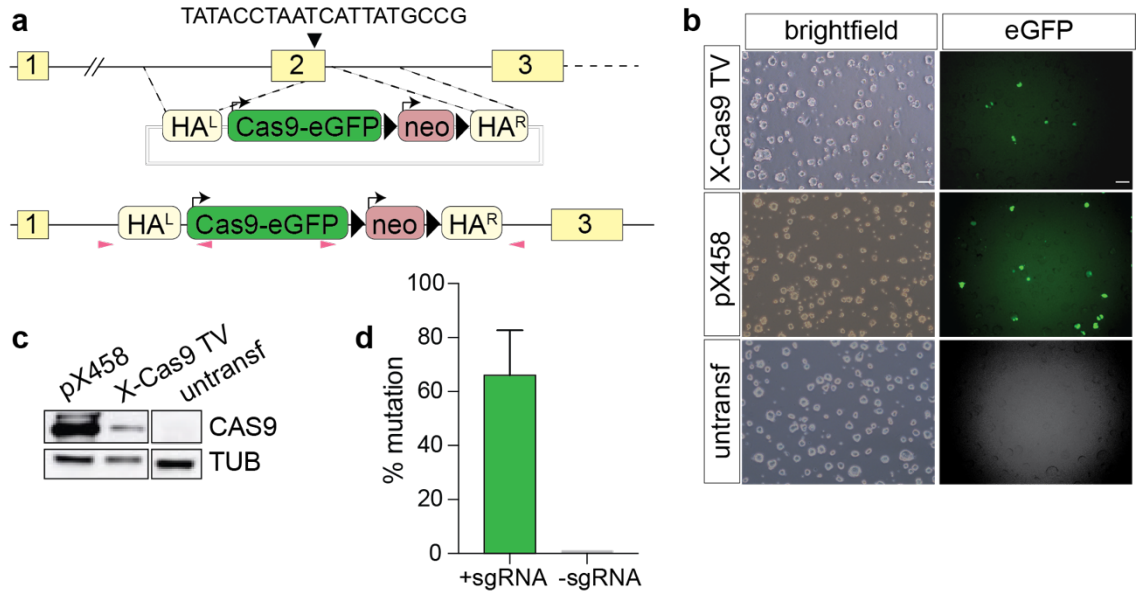

Figure 23. Generating the components for X-Cas9 HDR

(a) Targeting strategy for generating a X-Cas9 transgene by HDR. The TV encodes a Cas9-eGFP cassette driven by a CAG promoter and PGK promoter driven neomycin resistance cassette, flanked by LoxP sites. The TV contains two homology arms; 5' (HA left) and 3' (HA right) for HDR into *Hprt* exon 2. (b) The X-Cas9 TV or pX458 plasmid was transiently transfected in C57BL/6N ESCs to confirm eGFP expression by fluorescence microscopy. Scale bars=100µm (c) Western blots were carried out to assess Cas9 expression from protein extracted from transiently transfected ESCs. The western blot was performed by Valdone Maciulyte (Turner lab). (d) A CRISPR-Cas9 (pX458) plasmid containing a sgRNA targeting *Hprt* was transfected in C57BL/6N ESCs and mutation efficiency evaluated by the MiSeq and CrispRVariants analysis pipeline. The MiSeq library preparations were performed by the GeMS STP. The MiSeq sequencing was performed by the ASF. The CrispRVariants R pipeline was performed by Jasmin Zohren (Turner lab).

In conclusion, I generated the X-Cas9 TV encoding Cas9 and eGFP, for X-linked *Hprt* targeting by HDR. Co-transfection of the X-Cas9 TV and a *Hprt* sgRNA increases the efficiency of knock-in. Therefore, an sgRNA was confirmed to induce DNA DSBs at *Hprt* by the presence of CRISPR-Cas9 induced indel mutations.

6.2.2 Generating and characterising X-Cas9 embryonic stem cells

After confirming that the X-Cas9 TV expressed Cas9 and eGFP, I generated a stable X-Cas9 knock-in ESC line by HDR. Serum+LIF-maintained wildtype C57BL/6N ESCs were co-transfected with the X-Cas9 TV and *Hprt*-sgRNA plasmid. Two days post-transfection, neomycin was added to the medium to select for ESC clones that expressed the X-Cas9 plasmid. Neomycin selection was performed for eight days, followed by ESC colony picking and plating clones into individual wells of a 96-well plate (Figure 24a).

Individual colonies (n=48) were expanded and PCR-genotyped with a forward primer aligning to the endogenous *Hprt* locus and a reverse primer aligning to the CAG promoter in the transgene ("boundary PCR", Figure 23a pink arrow heads). Amplification of the 3.5kb product would only occur in ESCs with a targeted knock-in. The PCR product was amplified in nine ESC clones (n=9/48, 18.8%, Figure 24b). These ESC clones were called; a1, b6, c5, d3, d6, e6, f1, f5 and h5. The boundary PCR was also performed at the 3'end of the transgene for the nine positive ESC lines. I confirmed that in all nine clones there was successful amplification of the 3' boundary PCR product (4.1kb, Figure 24b). I therefore confirmed that the transgene was integrated at *Hprt* in nine ESC clones.

Boundary PCRs are extremely useful for assessing targeted integrations but do not determine whether the transgene has also integrated randomly elsewhere in the genome. Therefore, I next determined the transgene copy number in all nine X-Cas9 clones. Copy number greater than one inferred further transgene integrations. I performed copy number qPCR for eGFP and neomycin. Copy number was normalised to a reference gene *Tfrc* and a control homozygous *R26*-Cas9 sample, that has two copies of eGFP and neomycin.

In 78% of clones (n=7/9) there was less than two copies of eGFP and neomycin (a1, c5, d3, d6, e6, f1, h5, Figure 24c). Clone d3 appeared to have two copies of eGFP, but one copy of neomycin. Clone b6 had two copies of eGFP and neomycin (Figure 24c). Clone f5 had greater than six copies of eGFP and neomycin (Figure 24c). As a control, I also performed copy number qPCR for the X-linked gene, *Hprt*. I used a TaqMan probe to exons downstream of the HDR integration site, and therefore these exons should have been preserved. In all clones (n=9/9, 100%), *Hprt* was present in a single copy, as expected for XY ESCs (Figure 24c).

To confirm the copy number qPCR results, I also performed a digital droplet PCR (ddPCR). The eGFP copy number was normalised to *Tfrc*, a *R26*-Cas9 homozygous sample and a water control. The ddPCR results recapitulated the qPCR results. In the ddPCR, seven ESC lines (n=7/9, 78%) had one eGFP copy. These seven samples are the same samples that showed a single copy in the qPCR (a1, c5, d3, d6, e6, f1, f5). ESC line b6 again showed two copies. ESC line f5 again showed a significantly greater number of copies, exhibiting approximately five copies of the transgene (Figure 24d).

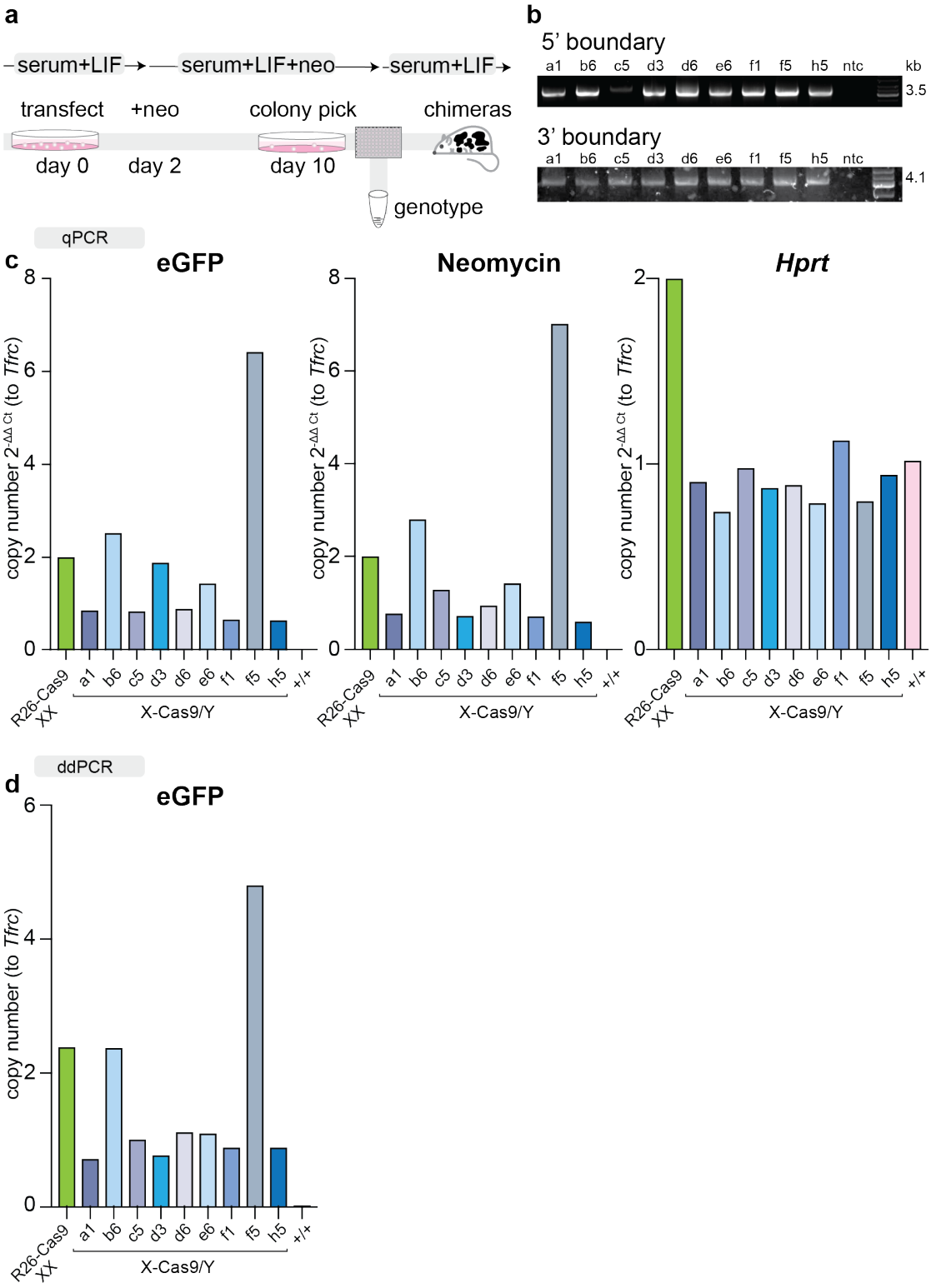

Figure 24. Characterising X-Cas9/Y ESCs

(a) Targeting strategy to integrate the Cas9-eGFP transgene onto the X chromosome in C57BL/6N ESCs. (c) PCR genotyping of surviving ESC clones. (d) Copy number qPCR for eGFP, neomycin and X-linked gene *Hprt*. Samples were normalised to the reference gene *Tfrc* in the *R26*-Cas9 (XX) homozygous control sample. (d) Copy number ddPCR for eGFP. Samples

were again normalised to the reference gene *Tfrc* in the *R26*-Cas9 (XX) homozygous control sample.

Next, I investigated if there were chromosomal aneuploidies in the X-Cas9 ESCs resulting from ESC culture. ESC aneuploidies may result in failure of the ESCs to contribute to the chimera and germline transmit. Genomic DNA was extracted from each ESC line, libraries prepared, followed by whole genome sequencing on the HiSeq 4000 at 0.1X coverage ("low-pass WGS"). The reads were aligned to the mouse genome (*Mm10*). Reads were binned into 500 kb regions, and log2 coverage from two autosomes was considered 0. The majority of samples showed consistent read average of 0 for all 18 autosomes, suggesting autosomal diploidy (Figure 25). In d3 and f5, some ESCs carried a duplication of chromosome 8, shown by a log2 of between 0 and 1 (arrows, Figure 25). The X chromosome was present at -1, suggesting a single X for all ESC lines (Figure 25). Although the Y chromosome reads are not in sufficient number to be visible on the plots, presence of a Y chromosome was confirmed by assessment of the raw reads, where Y-mapped reads are abundant (Table 2). Therefore, I confirmed that the majority of lines were diploid, and all lines were XY.

X-Cas9 sample ID	**Autosome reads**	**X-chromosome reads**	**Y-chromosome reads**
A1	16753566	534529	284071
B6	15003107	480932	253836
C5	15104788	477282	233640
D3	19060549	588769	316450
D6	21581887	688095	296232
E6	29019503	918110	485219
F1	26691797	854912	453075
F5	20747772	632858	338743
H5	24672453	787401	24404

Table 2. Sequencing reads (low-pass WGS) for X-Cas9 ESC clones

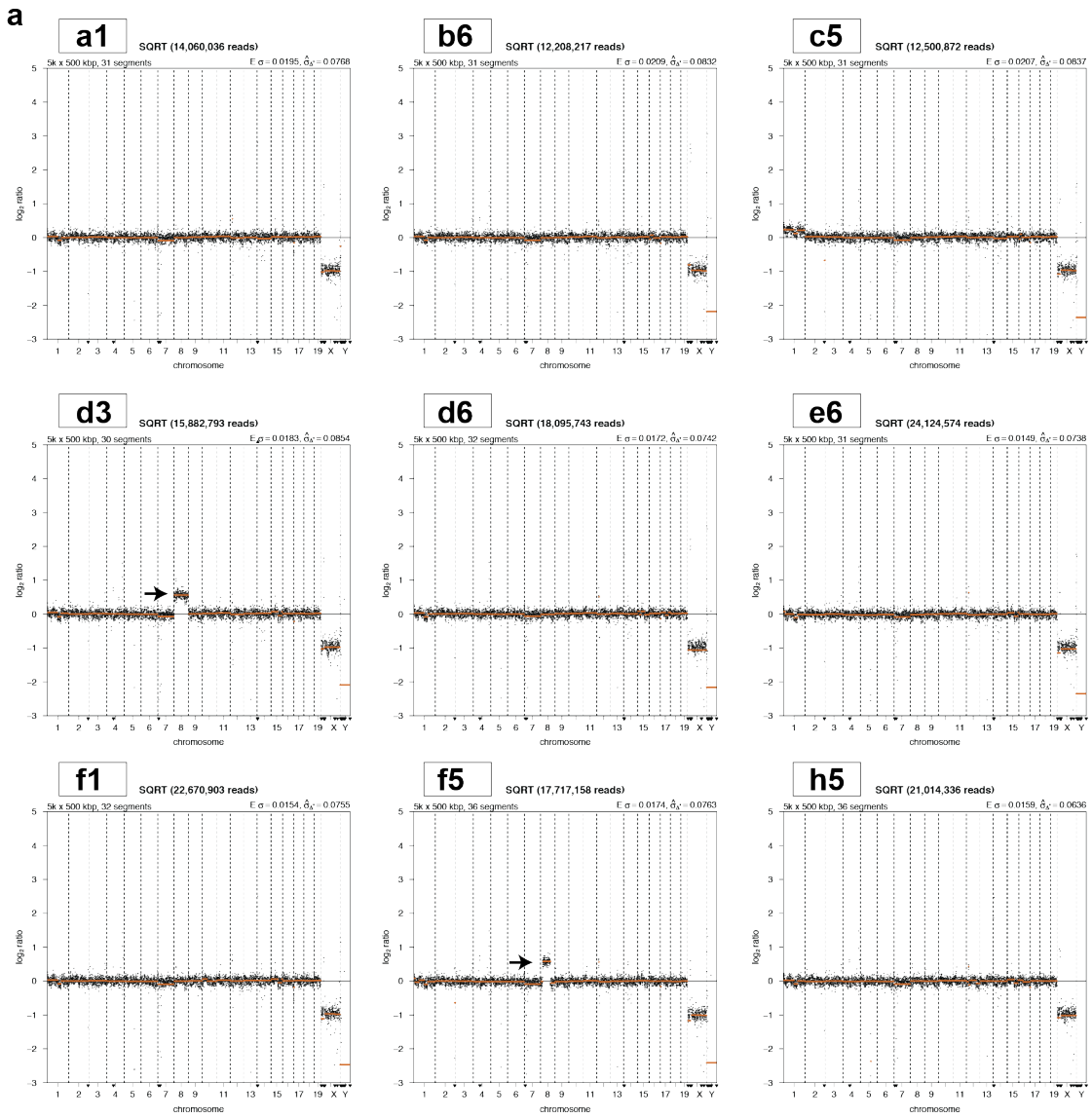

Figure 25. Low-pass whole genome sequencing of X-Cas9 clones for karyotyping

(a) HiSeq whole genome sequencing, 0.1X coverage, of X-Cas9 samples. Reads were binned into 500kbp regions and aligned to the mouse genome (mm10). The two autosomes for each mouse chromosome are considered as 0. The library preparations and sequencing were performed by the ASF. The R pipeline for analysis and to generate the plots was performed by Jasmin Zohren (Turner lab).

6.2.3 Assessing expression of X-Cas9

The majority of ESC clones were XY with a single copy of the X-linked Cas9 transgene by PCR-genotyping, copy analysis and low-pass WGS. I next determined if the ESCs expressed eGFP and Cas9. In all nine ESC lines, eGFP was expression was visible by fluorescence microscopy (Figure 26a).

Next, I performed qPCR for eGFP and *Hprt.* Gene expression was normalised to a housekeeping gene *Gapdh*, from wildtype C57BL/6J ESCs. I hypothesised that the X-Cas9/Y ESCs would express eGFP but have lost *Hprt* expression due to X-Cas9 transgene integration. In all nine X-Cas9/Y clones, eGFP was expressed while there was no eGFP expression in wildtype ESCs (Figure 26b). The strongest eGFP expression was from clone b6, and the weakest was from f1 (Figure 26b). The presence of multiple transgene copies in clones b6 and f5 did not significantly increase eGFP expression, compared with single copy clones. Confirming my hypothesis, there was a complete loss of *Hprt* expression in the X-Cas9/Y clones, compared with an XY wildtype sample and an XX *R26*-Cas9 sample (Figure 26b). The qPCR result therefore confirmed that the transgene had been successfully integrated into the *Hprt* locus, resulting in loss of *Hprt* expression.

I determined whether Cas9 expression was detectable at the protein level by performing a western blot for Cas9. The *R26*-Cas9 ESC line was used as a positive control, and ESCs and liver tissue from C57BL/6J mice were used as a negative control. Cas9 was expressed in all nine X-Cas9/Y ESC clones (n=9/9, 100%, Figure 26c), but expression was weaker than *R26*-Cas9 encoded Cas9. In summary, the X-Cas9/Y ESC clones express eGFP and Cas9.

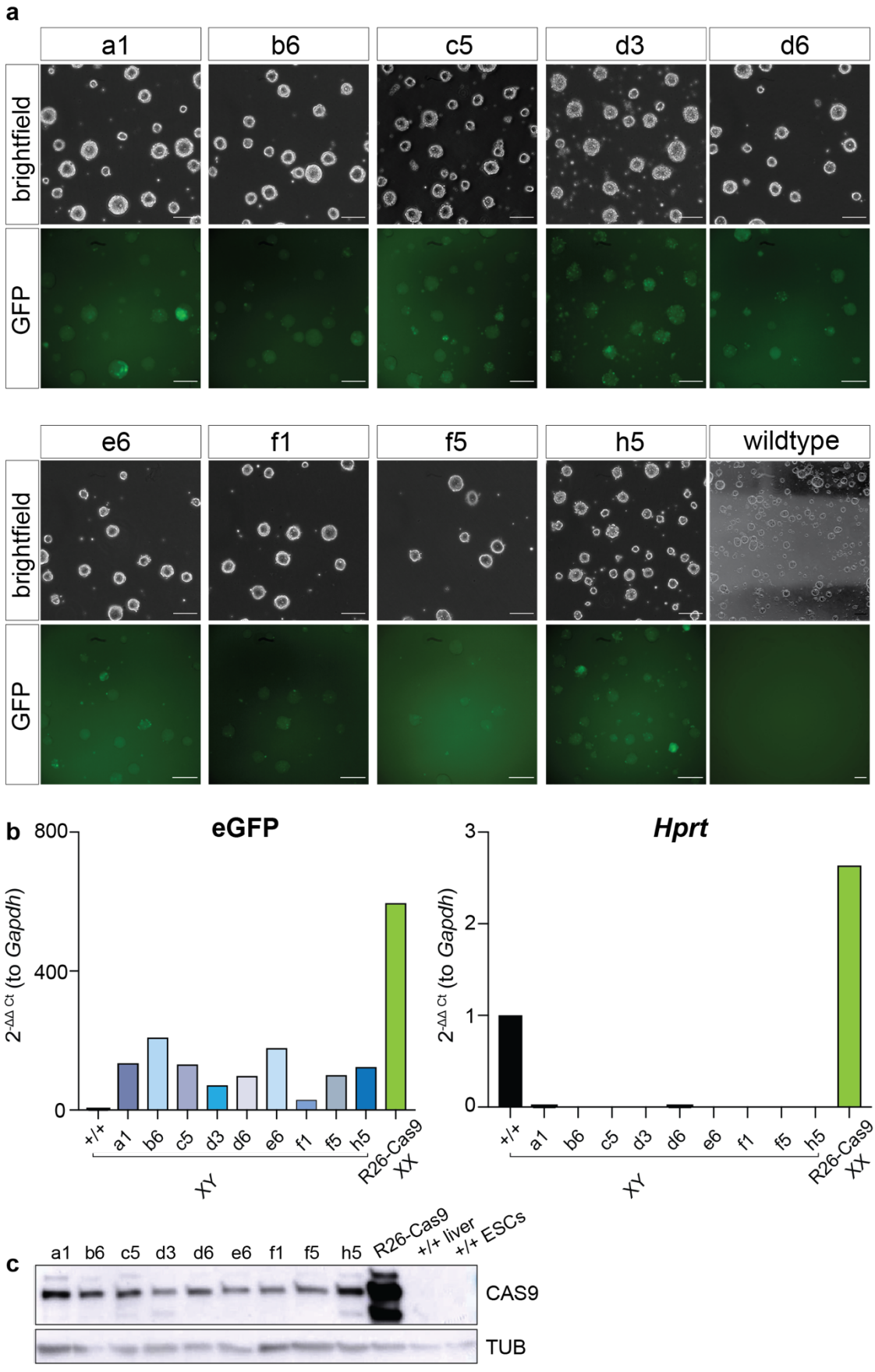
a
a1
b6
c5
d3
d6
brightfield
GFP
e6
f1
f5
h5
wildtype
b
eGFP
Hprt
800
400
0
3
2
1
2-ΔΔ Ct (to Gapdh)
+/+
a1
b6
c5
d3
d6
e6
f1
f5
h5
R26-Cas9
XY
XX
c
a1 b6 c5 d3 d6 e6 f1 f5 h5 R26-Cas9 +/+ liver +/+ ESCs
CAS9
TUB

Figure 26. X-Cas9 expression analysis

(a) Fluorescence microscopy images of X-Cas9 ESCs. (b) qPCR analysis of cDNA using TaqMan probes to quantify eGFP and *Hprt* expression, normalised to *Gapdh* in C57BL/6J. (c) Western blot of X-Cas9 ESCs using antibodies for CAS9 and TUBULIN. Comparing to *R26*-Cas9 ESCs and two C57BL/6J negative control samples; liver tissue and ESCs. The western blot was performed by Valdone Maciulyte (Turner lab).

6.2.4 Assessing the bi-component system *in vitro* using X-Cas9 ESCs

The X-Cas9/Y ESCs expressed eGFP and Cas9, therefore I investigated whether they would be functional in an *in vitro* bi-component system. I utilised the traffic light system of sorting transfected versus non-transfected ESCs by FACS, on reporter expression. I transfected the nine X-Cas9/Y ESC lines separately with the *Sry*-sgRNA pLethal plasmid. Transfected ESCs were mCherry+eGFP+, whilst non-transfected ESCs were eGFP+. The mCherry+eGFP+ ESCs were gated on mCherry fluorescence units greater than 10^3. The eGFP+ ESCs were gated on eGFP fluorescence units $10^{3.5}$ to $10^{4.5}$ (Figure 27). Post-sorting by FACS, the two populations of ESCs were lysed, *Sry* was PCR amplified for MiSeq sequencing and the reads analysed by the CrispRVariants pipeline (Lindsay et al, 2016).

In seven ESC lines (a1, b6, c5, d6, e6, f1, f5, h5) the mutation efficiency at *Sry* in mCherry+eGFP+ ESCs was greater than in the eGFP+ ESCs. The mean mutation efficiency in the mCherry+eGFP ESCs for these seven clones was 23.1%, whilst in the eGFP+ ESCs it was 1.7%. In the mCherry+eGFP+ ESCs from transfection of clone h5, the mutation efficiency was 10.7%, however in the eGFP+ ESCs, it was only slightly reduced, at 7.6% (Figure 27). In clone d3, a surprising converse result was seen, with the mutation efficiency in mCherry+eGFP+ ESCs at 9.6%, while in the eGFP+ ESCs it was 14.5% (Figure 27). Therefore, in seven ESC lines, the X-Cas9/Y ESCs are functional in a bi-component system, with the mutation at *Sry* being significantly higher when both CRISPR-Cas9 components are present, compared to Cas9 alone.

Considering the seven lines where the mutation efficiency was greater in the mCherry+eGFP+ ESCs, the non-Cas9-induced mutations in the eGFP+ ESCs averaged 1.7%. This pseudo-mutation efficiency is most likely attributable to PCR error during

amplification and library preparation, as in previous experiments using the bi-component system. The value of 1.7% is comparable to that seen in in Chapter 4 in the *R26*-Cas9 eGFP+ ESCs. Furthermore, this value may be attributable to the basal endogenous variability at the *Top1* locus. As in Chapter 4 and 5, the 1.7% mutation efficiency may also be attributable to contamination of mCherry+eGFP+ ESCs within the eGFP+ population, which will falsely increase the mutation efficiency of this population.

Curiously, in the bi-component experiment for clones h5 and d3, there is very little difference between the mutation efficiency in the mCherry+eGFP+ ESCs and the eGFP+ ESCs. I suspect that there was contamination between mCherry+eGFP+ and eGFP+ ESCs during the FACS, lysis and/or PCR amplification procedures, therefore these two populations are indistinguishable at the mutation efficiency level.

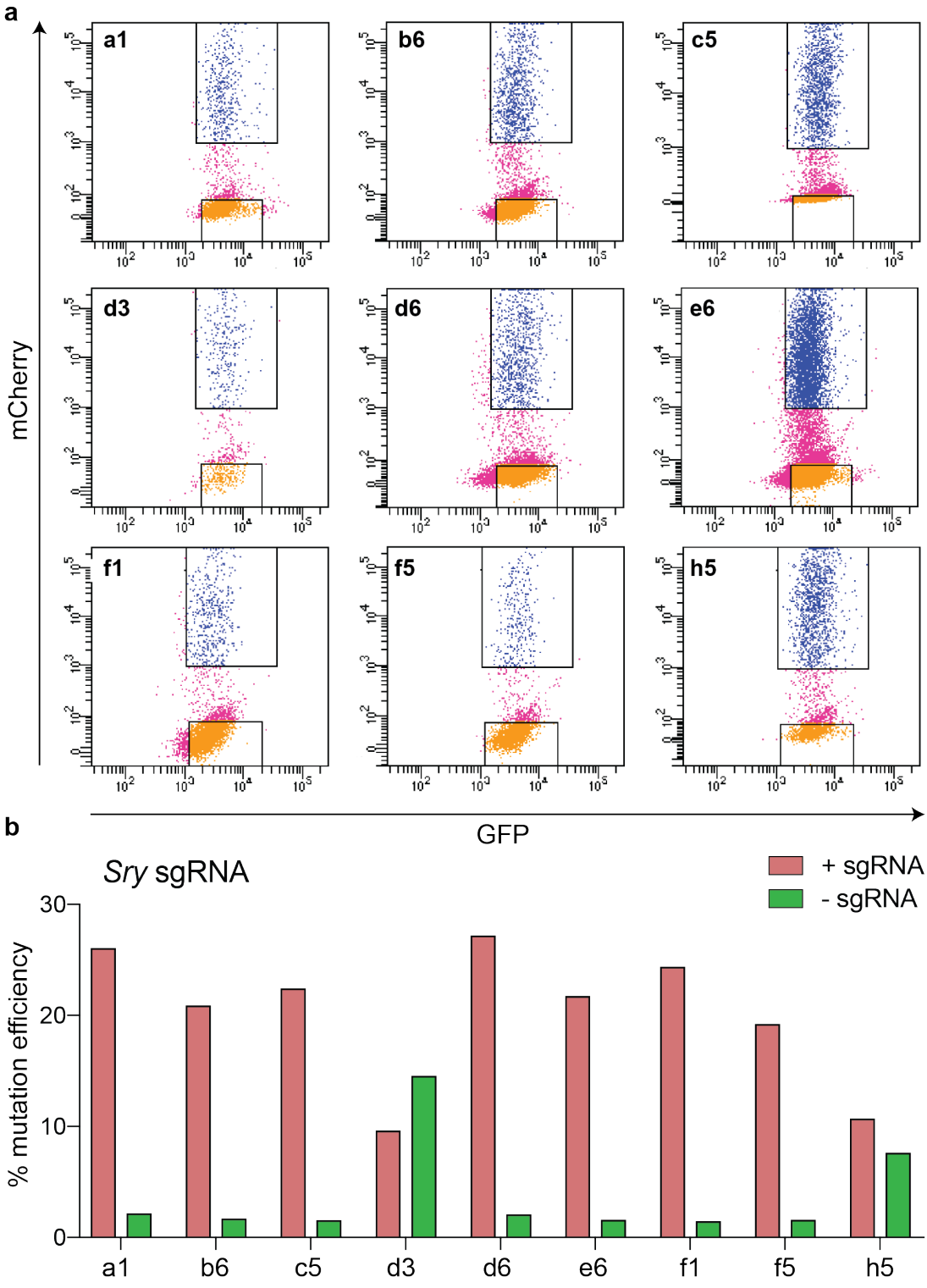

Figure 27. *Sry* sgRNA transfection and FACS/MiSeq in X-Cas9/Y ESCs

(a) FACS plots showing the gating for the mCherry+eGFP+ population (upper box in each plot) and the eGFP+ only population (lower box in each plot). The FACS was performed by the Flow Cytometry STP. (b) The ESC populations were lysed, PCR amplified, sequenced by MiSeq, and mutation efficiency quantified by CrispRVariants (Lindsay et al, 2016). The MiSeq library preparations were performed by the GeMS STP. The MiSeq sequencing was performed by the ASF. The CrispRVariants R pipeline was performed by Jasmin Zohren (Turner lab).

6.2.5 Generating X-Cas9 chimeras and assessing germline transmission

All of the X-Cas9/Y ESC clones expressed eGFP/Cas9 and the majority were functional in the bi-component system when the ESCs were transfected with a non-lethal sgRNA. I then used multiple X-Cas9/Y ESC lines to generate chimeras. The X-Cas9/Y ESCs were generated on a C57BL/6N genetic background and were therefore injected into albino C57BL/6J blastocysts, before surgical transfer into pseudopregnant females. When the pups were born I assessed the X-Cas9/Y ESC contribution by the percentage black coat colour.

The following lines were injected by the GeMS STP; a1, c5, d3, d6, e6, f5 and h5 (Figure 28a, Table 3). The surgical transfer was also performed by GeMS. Three clones did not produce any male chimeras. Of these three, clone c5 produced pups however they were all albino, showing no X-Cas9/Y ESC contribution (n=4/4 pups born, 100%). Clone d3 produced no pups. The lack of pups born was unsurprising given that this ESC line appeared to be trisomic for chromosome 8. Clone e6 produced one chimera with 50% black coat colour, but the chimera was female and was therefore unlikely to germline transmit the male XY ESC line. The remaining e6-produced pups were albino. The other four injected clones produced chimeras with varied percentage black coat colour. Clones a1, d6 and f5 all produced chimeras with a mean black coat colour above 50%. Clone h5 produced two male chimeras, but with low percentage contribution (20%, 25%, Figure 28a,b, Table 3). One high contribution male from injection of clone f5 had to be culled (red box around the data point, Figure 28a), and therefore could not be used to assess germline transmission. Representative images of some X-Cas9/Y chimeras produced by clones a1, d6 and h5 are shown in Figure 28b.

Clone	Number of embryos transferred	Pseudopregnant female transfers	Number of pregnancies	Chimeras/Total pups born (%)
A1	41	4	2	10/13 (77%)
C5	30	2	1	0/4 (0%)
D3	59	4	0	-
D6	75	5	2	5/8 (63%)
E6	60	5	2	1/9 (11%)
F5	30	2	1	2/2 (100%)
H5	25	2	1	2/3 (67%)

Table 3. Generating X-Cas9/Y chimeras

Male chimeras with high contribution from X-Cas9/Y ESCs were set up in matings with either albino C57BL/6J females, or autosomal constitutive Cre-expressing homozygous females Tg(Pgk1-Cre)1Lni, "PGK-Cre" (Lallemand et al., 1998). Chimeras generated from clones d6 and f5 were old enough to be set up in matings. One f5 male with a black coat contribution of 60% was set up in matings with albino C57BL/6J females. Two d6 chimeras with percentage black coat colour 95 and 92% were set up in matings with albino C57BL/6J females. One d6 chimera (95%) was set up in matings with PGK-Cre females. All four males were fertile, as assessed by successful pregnancies of females and live pups born (Figure 28c,d).

The F1 pups born from the d6 or f5 clone X-Cas9/Y chimeric males mated with albino C57BL/6J females were genotyped by coat colour and for the eGFP-containing transgene. Only one mating gave rise to black pups, male #3 (clone d6). This mating produced five pups born, four were male and were eGFP negative. There was one female pup who was eGFP positive (Figure 28e). In the X-Cas9/Y chimera mating to PGK-Cre females, there was a single litter born, with a total of four pups. None of the pups inherited the eGFP transgene (Figure 28e).

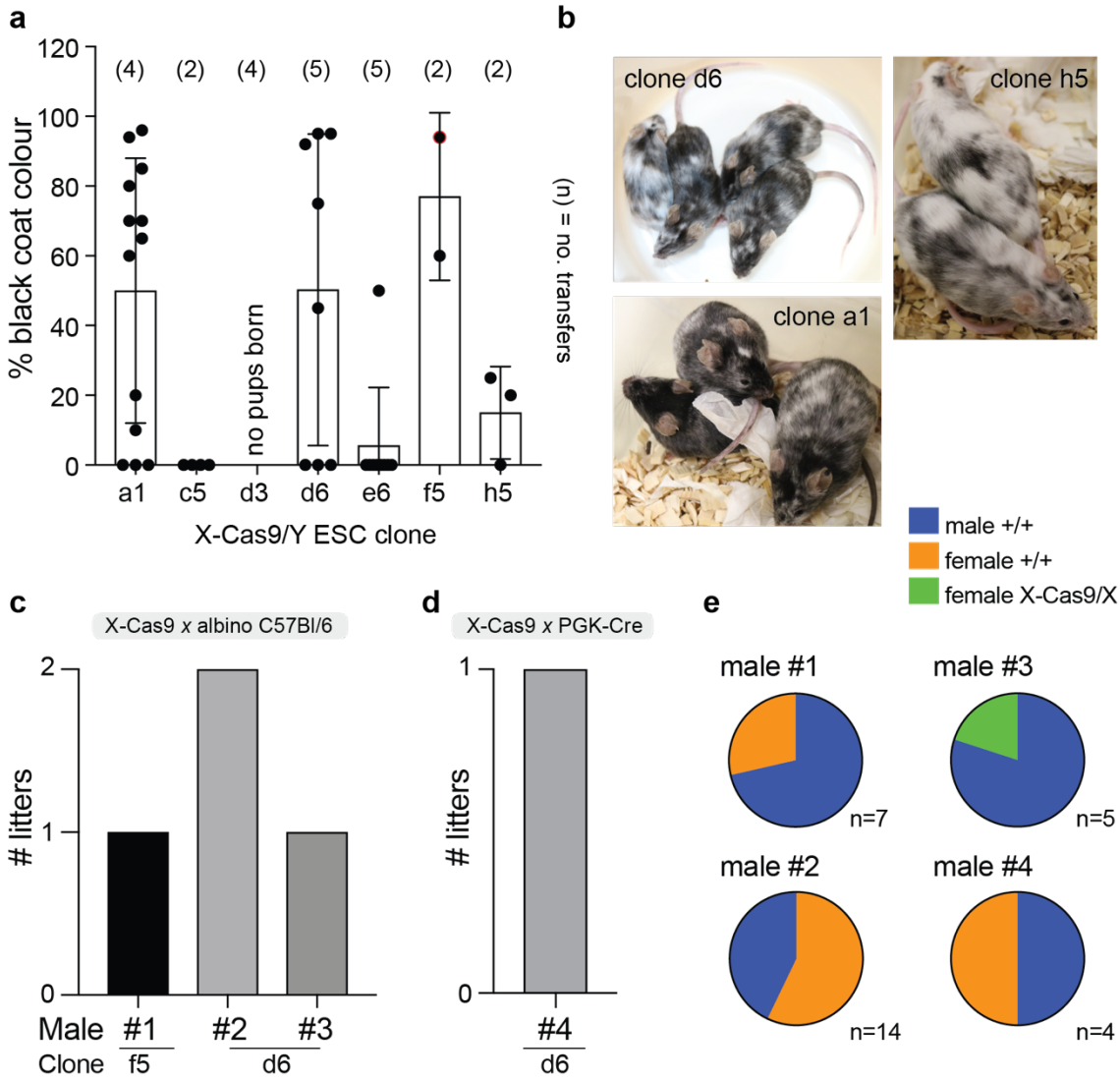

Figure 28. Generating X-Cas9/Y chimeras and germline transmission

(a) Percentage black coat colour contribution from X-Cas9/Y ESCs, on a C57BL/6N genetic background. Seven putative X-Cas9/Y ESC clones were injected. The ESC microinjection into embryos, and embryo surgical transfer was performed by the GeMS STP. (b) Images of X-Cas9/Y chimeras generated from clones d6, a1 and h5. (c) The number of litters produced from X-Cas9/Y chimeric males generated from clones f5 and d6, in matings with albino C57BL/6 females. (d) The number of litters produced from a single X-Cas9/Y male generated from clone d6 in matings with a PGK-Cre homozygous female. (e) Assessment of germline transmission of the X-Cas9 transgene by eGFP-genotyping.

In summary, there was a one female eGFP-positive pup born from a d6 X-Cas9/Y chimera mating with albino C57BL/6J females (Figure 28e, Figure 29a). To determine if the X-linked Cas9 transgene is expressed in this female, I performed a low-input western blot from an ear notch biopsy. Cas9 expression was also analysed from biopsies taken from the X-Cas9/Y chimeric father and an F1 male sibling that was eGFP negative.

In the X-Cas9/X F1 female and X-Cas9/Y chimeric father there was Cas9 expression, whilst there was no Cas9 expression in eGFP-negative male sibling (Figure 29b).

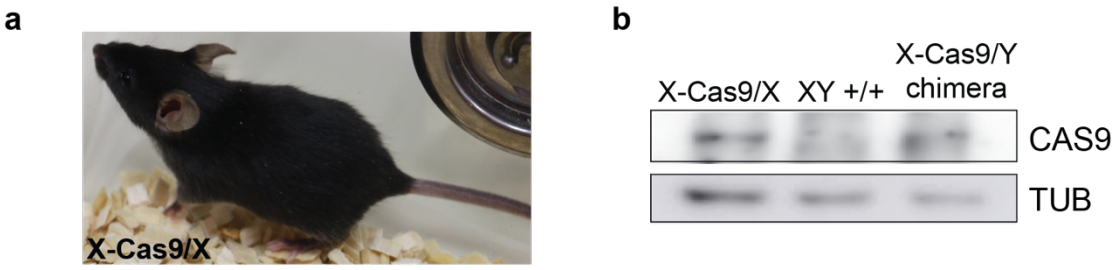

Figure 29. Cas9 expression in a X-Cas9/X F1 generation female

(a) Female X-Cas9/X F1 generation female with black coat colour from matings of X-Cas9/Y chimera to albino C57BL/6J females. (b) Western blot for Cas9 expression. The western blot was performed by Valdone Maciulyte (Turner lab).

I generated an X-Cas9 stable transgenic mouse line that germline transmitted successfully to the F1 generation. The Cas9 transgene was inherited uniquely by female offspring from the chimeric father generated from clone d6, confirming that the transgene was X-linked. Cas9 expression was detected both in ESCs and in low-input ear biopsy tissue. However unfortunately due to time constraints, the X-Cas9 mouse line could not be tested in the *in vivo* bi-component system by mating to the sgRNA mouse line.

6.2.6 CRISPR-Cas9 HDR components for Y-Cas9 HDR

In the complementary strategy, where the Cas9 transgene is only inherited by sons, the transgene is Y chromosome-linked. I generated a Y-Cas9 TV and sgRNA targeting an intergenic Y chromosome region. The Y-Cas9 TV also encoded eGFP, where Cas9 and eGFP were driven by a CAG promoter.

I generated a TV that targeted an intergenic Y chromosome region downstream of *Uty* and *Eif2s3y*, two highly expressed, constitutively active genes. The constitutive expression of *Uty* and *Eif2s3y* suggested this was a euchromatic region. The homology arms were amplified from C57BL/6J gDNA, aligning to the intergenic Y chromosome region. The 5' homology arm (0.6kb; HA^L) was inserted into pX458 using unique

restriction sites XbaI and KpnI. The 3' homology arm (0.6kb; HA^R) was inserted into unique restriction site NarI via Gibson Assembly cloning methods (Figure 30a). There was 52bp of intergenic sequence between the two homology arms. A PGK promoter-driven neomycin cassette, flanked by LoxP sites, was inserted into the TV by Gibson Assembly cloning.

Alongside earlier described experiments assessing expression of Cas9 and eGFP from the X-Cas9 TV by transient ESC transfection, contemporaneously I confirmed expression of Cas9 and eGFP from the Y-Cas9 TV. Wildtype ESCs were transfected with either the backbone pX458 plasmid, the Y-Cas9 TV or were left as an untransfected control. In the ESCs transfected with the Y-Cas9 TV, eGFP was expressed in some cells (Figure 30b), similarly to the X-Cas9 TV and pX458 plasmid. To confirm that the Y-Cas9 TV expressed Cas9, I extracted bulk protein from the ESCs and western blots for Cas9 were performed. Results showed that the ESCs transfected with the Y-Cas9 TV expressed Cas9, similarly to pX458 (Figure 30c). There was no Cas9 expression in the untransfected controls, as expected. Therefore, I confirmed that the Y-Cas9 TV expressed eGFP and Cas9.

I also generated a plasmid expressing an sgRNA targeting the Y-chromosome intergenic region within the 52bp intervening sequence between the two homology arms. The 20 nucleotide oligonucleotides were cloned into pX458 and confirmed by Sanger sequencing. I transfected the Y-chromosome sgRNA pX458 plasmid into wildtype C57BL/6N ESCs. The transfected and non-transfected ESCs were sorted by FACS on eGFP reporter expression. The eGFP+ and eGFP- ESCs were lysed and the Y-intergenic target region amplified by PCR, sequenced by MiSeq and analysed by the CrispRVariants pipeline (Lindsay et al, 2016). The mean mutation efficiency at the Y-chromosome intergenic position was 17.9% in the eGFP+ ESCs (Figure 30d), highlighting the ability of the sgRNA to produce a DNA DSBs at the target region.

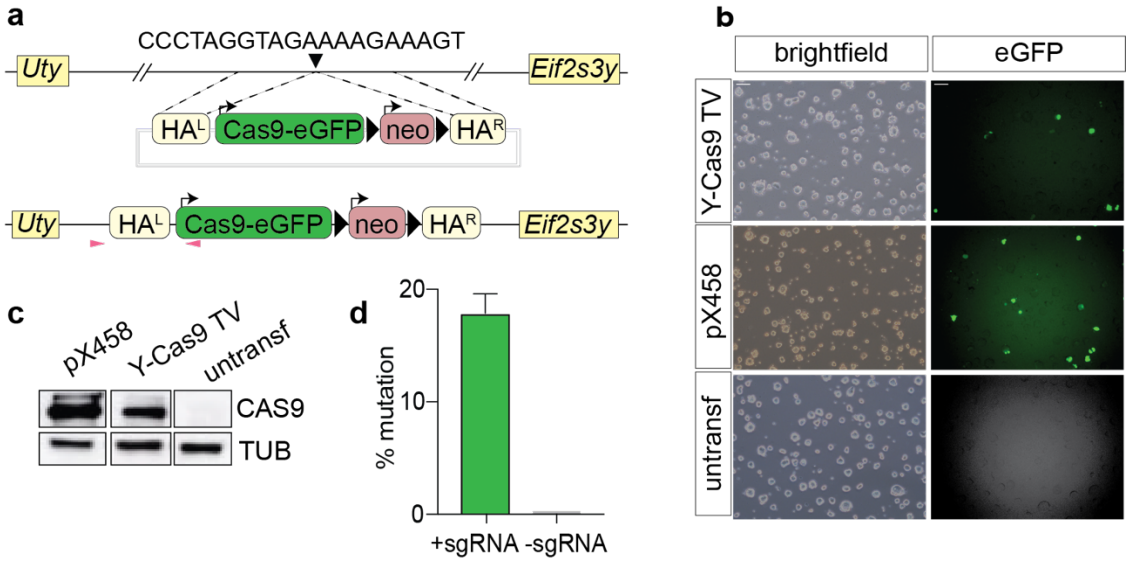

Figure 30. Generating the components for Y-Cas9 HDR

(a) Targeting strategy for HDR at the Y chromosome intergenic region. The transgene to be inserted contains Cas9-eGFP under a constitutive CAG promoter and neomycin under a constitutive PGK promoter, flanked by LoxP sites. The transgene is bordered by two homology arms; 5' (HA left) and 3' (HA right) for homology directed recombination at the target locus. (b) The plasmid targeting vector was transiently transfected in wildtype C57BL/6N embryonic stem cells to confirm eGFP expression by fluorescence microscopy. Scale bars=100µm (c) The transiently transfected embryonic stem cells were collected and bulk protein extracted for western blot. Antibodies against Cas9 and Tubulin were used for detecting gene expression at the protein level. The western blot was performed by Valdone Maciulyte (Turner lab). (d) A CRISPR-Cas9 plasmid (pX330) containing an sgRNA targeting the Y chromosome was tested to ensure cleavage at the target position. Mutation efficiency was evaluated by PCR amplification of the target region, sequencing by MiSeq and evaluation by CrispRVariants. The MiSeq library preparations were performed by the GeMS STP. The MiSeq sequencing was performed by the ASF. The CrispRVariants R pipeline was performed by Jasmin Zohren (Turner lab).

Overall, this data shows that DNA DSBs were generated by CRISPR-Cas9 at the intergenic Y chromosome region, downstream of *Uty* and *Eif2s3y*. I theorised that induction of the DNA DSB increases HDR efficiency at the Y chromosome for Y-Cas9 HDR.

6.2.7 Generating and characterising Y Cas9 embryonic stem cells

Next, I generated the Y-Cas9 knock-in via CRISPR/Cas9 induced HDR in C57BL/6N ESCs. Serum+LIF-maintained ESCs were co-transfected with the Y-Cas9 TV and sgRNA plasmid. Two days post-transfection, neomycin was added to the cell culture medium for eight days. A total of 384 colonies were picked into individual wells of a 96-

well plate, DNA extracted and PCR-genotyped (Figure 31a). To PCR-genotype, I performed a 5' boundary PCR (primers shown by arrowheads, Figure 30a). Of 384 colonies, one clone was boundary PCR positive, "1b5" (Figure 31b). The PCR product was Sanger sequenced and results confirmed an integration at the target site, by the presence of endogenous Y-sequence adjacent to homology arm sequence, immediately followed by transgene sequence. There were no indels around the 5' integration site. Conversely however, the 3' boundary PCR consistently failed, despite extensive trouble shooting.

To determine whether there were other random transgene integrations elsewhere in the genome in ESC line 1b5, a copy number qPCR was performed. The eGFP and neomycin copy number was normalised to a reference gene *Tfrc* in *R26*-Cas9 homozygous sample (eGFP and neomycin present in two copies). In 1b5, the copy number of eGFP and neomycin was one, suggesting they are present in a single copy (Figure 31c). To confirm the qPCR copy number analysis, a Southern blot was performed using a neomycin probe and the *R26*-Cas9 ESC line as a positive control for probe hybridisation. In 1b5 there was a single band at the expected height when probing from both the 5' (8.5kb) and 3' (14.5kb) direction (Figure 31d). Overall, these data confirmed that there was a single integration of the Cas9-eGFP transgene at the Y-intergenic region.

To determine if there were any chromosomal aneuploidies in the Y-Cas9 ESC line, low-pass WGS was performed. The results showed that the autosomes were diploid. The X chromosome was present in a single copy, and reads aligned to the Y chromosome (not visible in the plot, read number listed in Table 4), confirming that 1b5 was XY (Figure 31e).

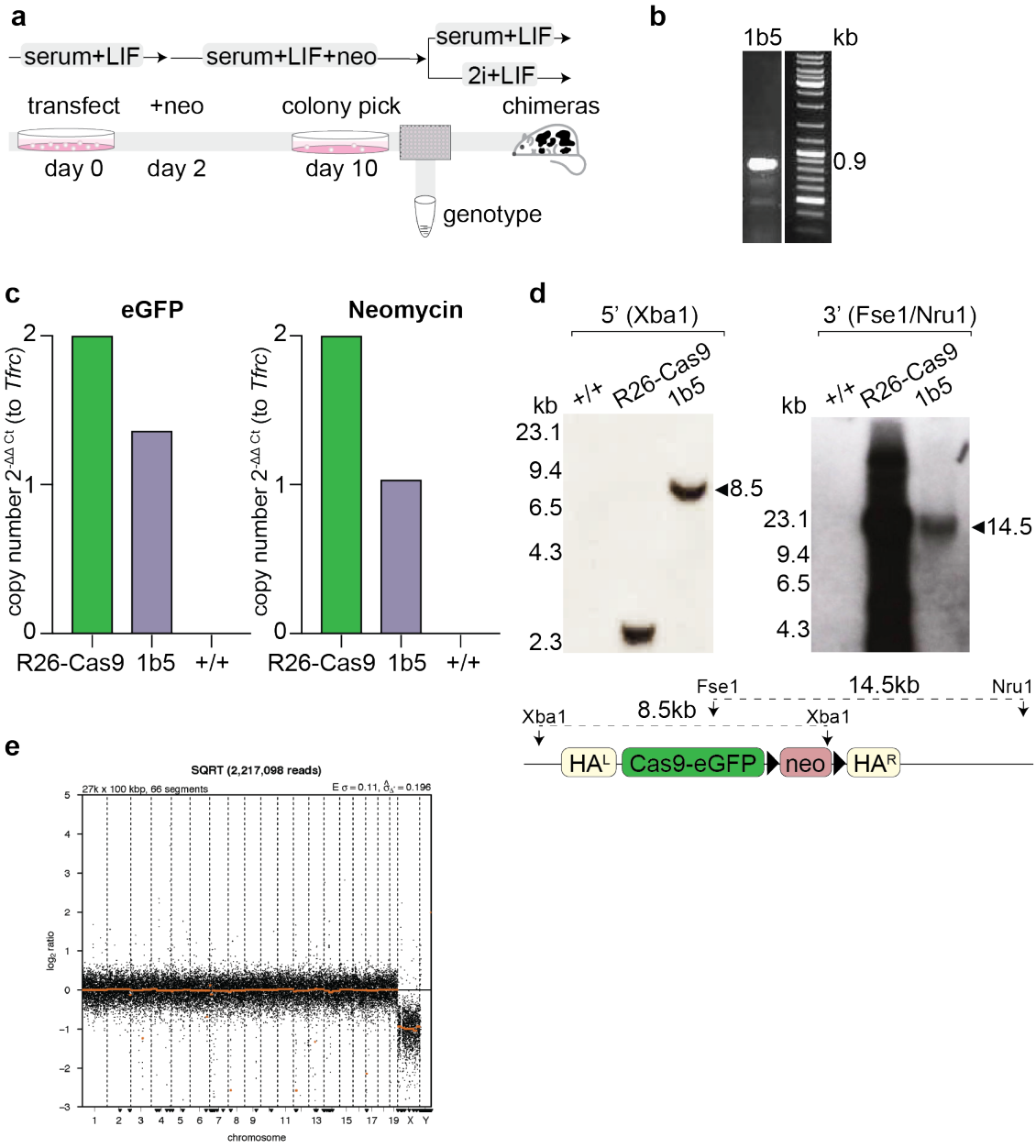

Figure 31. Generating a Y-Cas9 knock-in embryonic stem cell line

(a) Strategy for generating a knock-in by HDR in C57BL/6N embryonic stem cells. Cells were transfected with CRISPR-Cas9 components and a Y-Cas9 plasmid targeting vector containing a neomycin resistance cassette. After 2 days post-transfection cells were selected by addition of neomycin antibiotic into the 2i+LIF medium. After 8 days selection, surviving colonies were picked into individual wells of a 96-well plate and expanded for genotyping. (b) PCR genotyping of Y-Cas9 clones using boundary PCR. (c) Copy number qPCR for eGFP and neomycin, normalised to *Tfrc* in *R26*-Cas9 homozygous sample. (d) Southern blot (5' and 3') of clone 1b5 to determine the number of integrations of the transgene, probing for neomycin. The Southern blots were performed by Valdone Maciulyte (Turner lab). (e) Assessing chromosomal aneuploidy by DNA extraction from clone 1b5, low-pass WGS and bioinformatic analysis. The library preparations and sequencing were performed by the ASF. The R pipeline for analysis and to generate the plots was performed by Jasmin Zohren (Turner lab).

Y-Cas9 sample ID	Autosome reads	X chromosome reads	Y chromosome reads
1b5	2828894	89030	49025

Table 4. Sequencing reads (low-pass WGS) for the Y-Cas9 ESC clone.

Overall, these results showed that X/Y-Cas9 clone 1b5 contained a single integration of the Cas9-eGFP transgene at the target Y chromosome intergenic region, and was therefore suitable for blastocyst injection to generate chimeras.

6.2.8 Producing Y Cas9 chimeras and assessing germline transmission

After determining that the Cas9-eGFP transgene was a single copy integration at the expected position on the Y chromosome, the GeMS STP generated chimeras by blastocyst injection of 1b5. Passage 3 serum+LIF-maintained 1b5 ESCs were injected into albino C57BL/6J blastocysts and GeMS surgically transferred the injected embryos into pseudopregnant females. I assessed the pups born for black coat colour chimerism. A total of 27 pups were born and 18 (67%) had black coat colour contribution (Table 5, Figure 32a). Of the pups that showed any X/Y-Cas9 contribution, the mean black coat colour percentage was 45%. The methylation status of 2i+LIF-maintained ESCs is considered to be more representative of the epiblast and therefore 2i+LIF ESCs may contribute more highly to chimeras (Mulas et al., 2019b). Therefore, passage four 1b5 X/Y-Cas9 ESCs were transferred into 2i+LIF for three passages (Figure 31a). After 2i+LIF culture, the ESCs were injected into albino C57BL/6J blastocysts by the GeMS STP. Of 28 pups born, 71% (n=20/28) had black coat colour contribution from X/Y-Cas9 ESCs. The mean black coat colour contribution was 62% (Table 5, Figure 32a,b) and one pup had 100% black coat (can be seen in Figure 32b). These results suggested that short maintenance of ESCs in 2i+LIF may increase ESC-contribution to the chimera.

Clone	Number of embryos transferred	Pseudopregnant female transfers	Number of successful pregnancies	Chimeras/Total pups born (%)
1b5 (serum+LIF)	80	4	3	18/27 (67%)
1b5 (2i+LIF)	96	8	4	20/28 (71%)

Table 5. Generating the X/Y-Cas9 chimeras

Next, I assessed Y-Cas9 germline transmission from the chimeras. The Y-Cas9 transgene was integrated 40kb and 10kb downstream of *Uty* and *Eif2s3y,* respectively. *Eif2s3y* knock-out results in male infertility (Matsubara et al., 2015) and therefore disruption to *Eif2s3y* from transgene insertion may have the same phenotype. Therefore, I first determined if the high contribution chimeras were fertile by successful mating with wildtype females. Four chimeric males were set up in matings with wildtype females. Two of the males were produced from serum+LIF 1b5 ESCs (male #1 and #2), and two were produced from 2i+LIF-transferred ESCs (male #3 and #4). Males #1 and #2 were both 75% black. Male #3 was 90% black, and male #4 was 100%. Three of the males produced multiple litters by wildtype females; #2, #3 and #4. Therefore, the fertility of these three males was not impaired by insertion of the transgene (Figure 32c).

Once I had confirmed the chimeric males were fertile, I ear biopsy-genotyped the offspring to determine if the transgene was germline transmitted. The male offspring should uniquely inherit the transgene; however all pups were genotyped to confirm this. Of the three fertile males, all germline transmitted the Cas9-eGFP transgene to offspring (Figure 32d). From male #2 (75% chimeric, serum+LIF), only 11% male pups born was X/Y-Cas9 (n=1/9). From male #3 (90% chimeric, 2i+LIF), 58% male pups were X/Y-Cas9 (n=7/12), while from male #4, (100% chimeric, 2i+LIF), 90% male pups were X/Y-Cas9 (n=9/10; Figure 32d). I also confirmed there were no eGFP positive female offspring born (n=0/31 total females born, 0%; Figure 32d).

The Y-Cas9 F1 males were assessed for fertility by testis weight measurement. Both testes were removed by dissection from culled males, weighed and the two values averaged. The mean testis weight was calculated from seven X/Y-Cas9 males and five wildtype C57BL/6J males (+/+). The mean testis weight for wildtype males was 98mg, while for the X/Y-Cas9 it was 103mg (Figure 32e).

To confirm Y-linked gene expression from F1 generation males, I performed qPCR for *Uty* and *Eif2s3y*. Y-linked gene expression was normalised to *Gapdh* from a wildtype male. In the X/Y-Cas9 male, expression levels of *Uty* and *Eif2s3y* were similar to that of a C57BL/6J wildtype male. There was no expression of *Uty* and *Eif2s3y* in control female tissue (Figure 32f). Expression of the X-linked gene *Hprt* was also analysed as a housekeeping gene control. *Hprt* was expressed in wildtype male and female, and X/Y-Cas9 tissue (Figure 32f). Overall, this data shows that expression of Y-linked genes *Uty* and *Eif2s3y* were unimpaired in the X/Y-Cas9 male F1 generation.

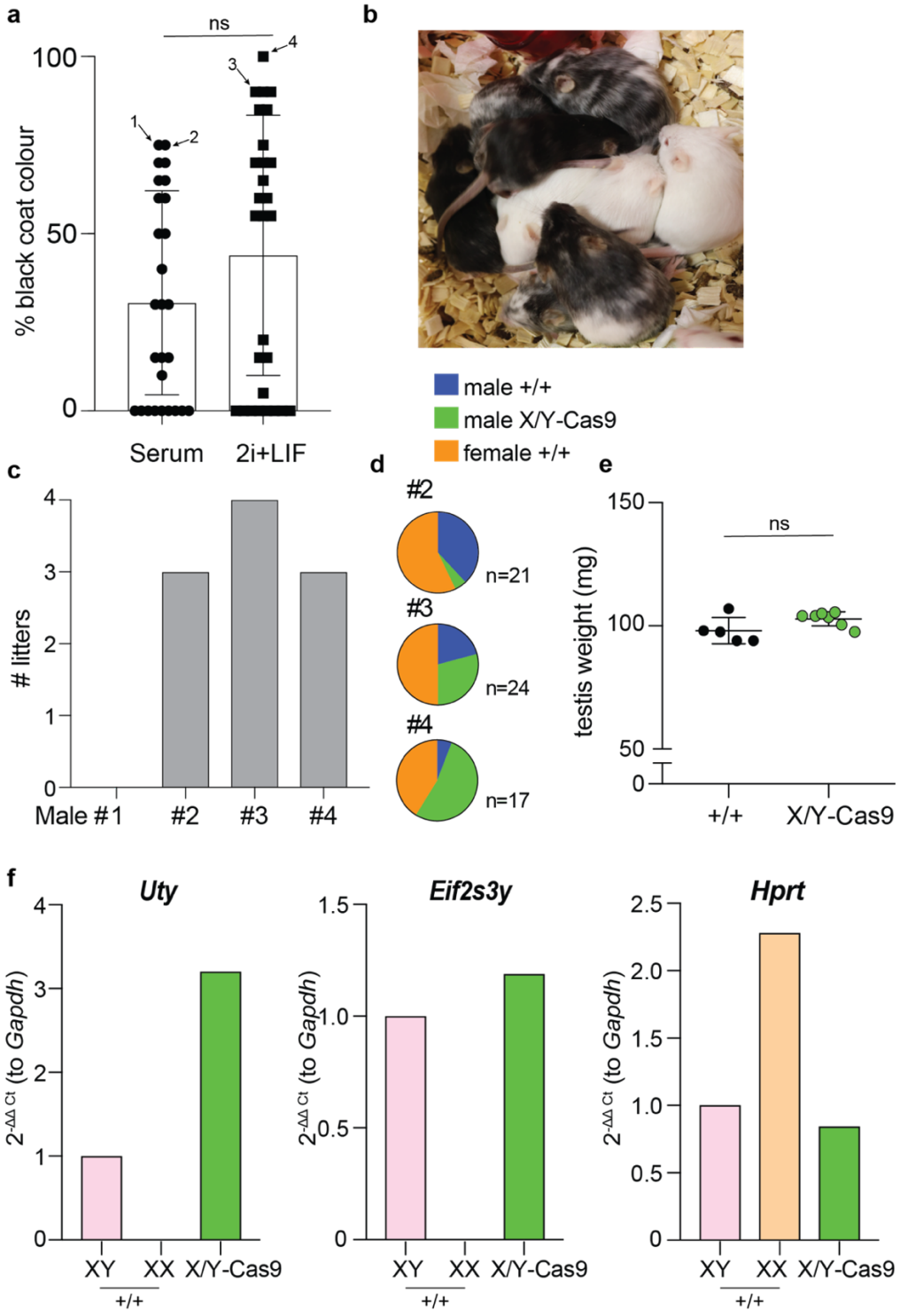

Figure 32. Generating Y-Cas9 chimeras, germline transmission and assessing fertility

(a) Coat colour contribution from black X/Y-Cas9 ESCs after culture in serum+LIF or 2i+LIF conditions. (b) Example chimeras generated from injection with clone 1b5. The ESC microinjection into embryos, and embryo surgical transfer was performed by the GeMS STP. (c) Litters produced after mating with wildtype females. (d) Assessment of germline transmission of the Y-Cas9 transgene. (e) Testis weights of C57BL/6J wildtype males (+/+) versus X/Y-Cas9 males. (f) qPCR analysis of Y chromosome linked genes, *Uty* and *Eif2s3y* and X-linked gene *Hprt*.

6.2.9 Investigating Cas9 expression by Cre recombinase removal of neomycin

The Y-Cas9 transgene was stably integrated into the Y-chromosome intergenic region and was successfully germline transmitted. I next assessed Cas9 and eGFP expression. The PGK promoter driving neomycin expression could have local silencing effect on the transgene. Therefore, I assessed Cas9 and eGFP expression from the Y-Cas9 transgene both with PGK-neo and post-Cre excision.

Protein from an X/Y-Cas9 male, wildtype male and *R26*-Cas9 male was extracted and western blots performed to assess Cas9 expression. In the *R26*-Cas9 positive control, Cas9 was expressed. Conversely, in the wildtype and X/Y-Cas9 samples, there was no Cas9 expression (Figure 33a). To determine if the Cas9 deficiency was *in vivo* specific, the X/Y-Cas9 1b5 ESCs were also tested by western blot. Similarly to the X/Y-Cas9 tissue, there was no Cas9 expression (Figure 33b). The western blot results show that the X/Y-Cas9 transgene is not expressed *in vivo* or *in vitro*.

To excise the floxed PGK-neomycin cassette *in vivo*, X/Y-Cas9 males were set up in matings with X-Cre hemizygous females (X-Cre/X). Pups were genotyped for Cre and the transgene. Two X-Cre/Y-Cas9 males, 3.1a and 3.1c, were used for future experiments. To confirm that the floxed-PGK cassette was excised I PCR-genotyped the two X-Cre/Y-Cas9 males. I utilised PCR primers that align to the transgene, 5' of the PGK-neomycin cassette, and to the homology arm, 3' of the cassette (orange arrows, Figure 33c). If the PGK-neomycin is present, the amplicon is 2.2kb. If the PGK-neomycin has been lost, the amplicon had a size shift to 0.85kb (orange arrows, Figure 33c). The two X-Cre/Y-Cas9 males showed the amplicon size shift compared to X/Y-Cas9 males (Figure 33d), confirming the floxed-neomycin had been excised.

I tested Cas9 and eGFP expression post-Cre recombinase excision of the floxed-PGK-neomycin at the RNA and protein level, by performing qPCR and western blot. In the western blot I included the positive control *R26*-Cas9, negative control C57BL/6J and parental X/Y-Cas9 tissue. Results showed that there was no Cas9 expression in either X-

Cre/Y-Cas9 males, similar to the X/Y-Cas9 father (Figure 33e). In the qPCR, expression of Y-linked genes *Uty*/*Eif2s3y* and transgenic eGFP was normalised to *Gapdh* in wildtype tissue. In the X-Cre/Y-Cas9 males, there was expression of *Uty* and *Eif2s3y,* however eGFP was not expressed (Figure 33f).

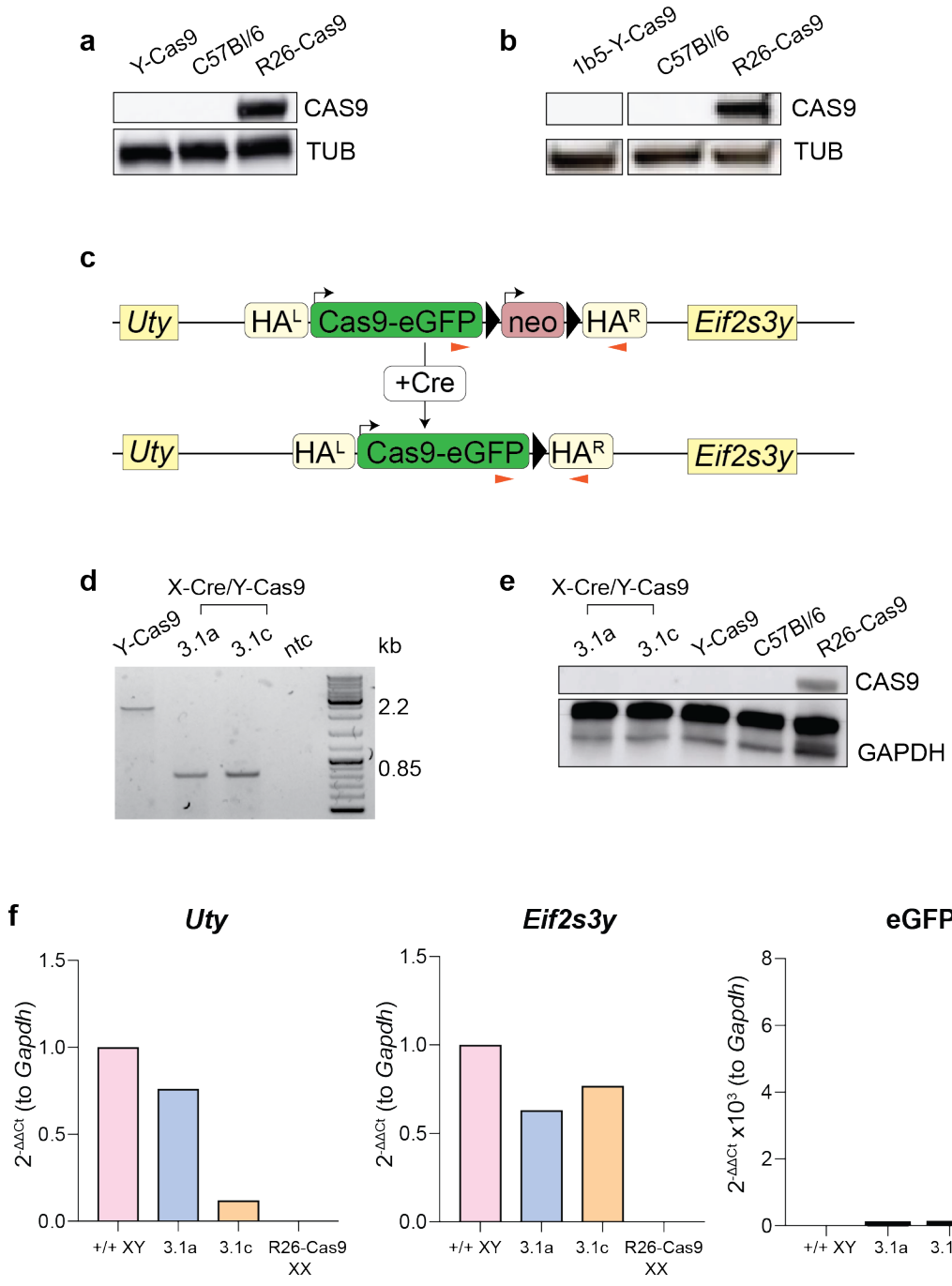

Figure 33. Transgene expression in X/Y-Cas9 and X-Cre/Y-Cas9 males

(a) Western blot to assess expression of Cas9 in the Y-Cas9 mouse. (b) Western blot to assess expression of Cas9 in the 1b5 Y-Cas9 ESC line. All western blots were performed by Valdone Maciulyte (Turner lab). (c) Cre recombinase removal of the floxed-PGK-neomycin cassette *in vivo* by crossing a Y-Cas9 male with a X-Cre female mouse. (d) PCR genotyping highlighting removal of the floxed-PGK-neomycin cassette *in vivo*. (e) Western blot to assess expression of Cas9 in X-Cre/Y-Cas9 male mice after removal of the floxed-PGK-neomycin cassette. (f) qPCR analysis of Y chromosome genes *Uty* and *Eif2s3y*, and transgene eGFP, in X-Cre/Y-Cas9 males.

It was possible that the lack of transgene expression in the X-Cre/Y-Cas9 males was due to mosaicism of floxed-neomycin excision. To assure that this was not the case, I set up X-Cre/Y-Cas9 males in matings with wildtype females (Figure 34a). Male pups born were genotyped for the Y-Cas9 and X-Cre alleles. All female offspring were X-Cre/X (n=7) while all males were X/Y-Cas9 (-neo; n=8, Figure 34b). Two X/Y-Cas9(-neo) males, 6.1a and 6.1b, were carried forward for future experiments. I confirmed that there was a complete loss of the floxed-neomycin cassette by PCR genotyping, determined by the presence of the size-shifted amplicon. In both X/Y-Cas9(-neo) males (hereafter "X/Y-Cas9"), the neomycin cassette had been excised (Figure 34c). I also confirmed that the eGFP had remained in a single copy by performing copy number ddPCR, normalising to *R26*-Cas9 homozygous *Tfrc*. The ddPCR results showed that the eGFP was single copy in the X/Y-Cas9 males (Figure 34d).

I determined whether complete removal of the PGK-neomycin induced eGFP expression by qPCR. Expression of transgenic eGFP and X-linked gene *Hprt* was normalised to *Gapdh* expression in wildtype tissue. Two different tissues were tested for each X/Y-Cas9 male. The qPCR results showed that there was low, but detectable eGFP expression in both tissue types in the X/Y-Cas9 males (Figure 34e). The expression of X-linked gene *Hprt* was comparable between control samples and X/Y-Cas9 samples (Figure 34f).

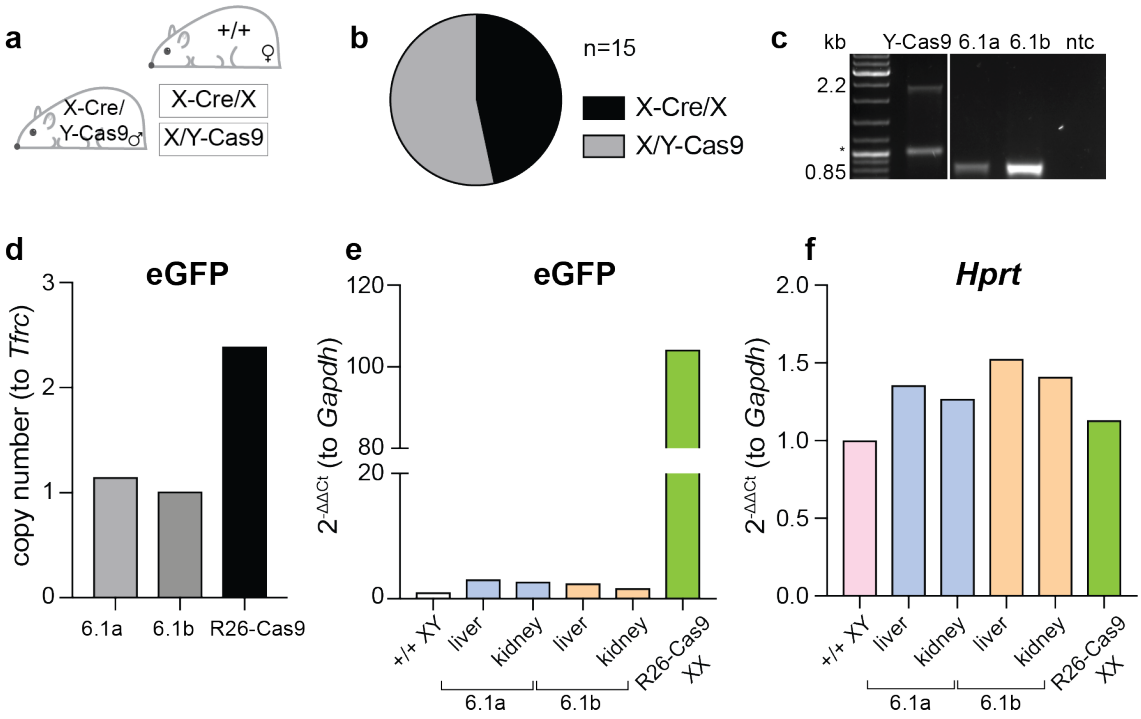

Figure 34. eGFP expression in X/Y-Cas9 males from X-Cre/Y-Cas9 parents

(a) X-Cre/Y-Cas9 males were set up for matings with wildtype C57BL/6J females (+/+) to ensure all Y-Cas9 offspring were non-mosaic for removal of the neomycin cassette. (b) Offspring genotypes (n=15 pups): X-Cre/X females (n=7) and X/Y-Cas9 males (n=8). (c) Two X/Y-Cas9 males were genotyped for loss of the neomycin cassette. (d) Confirmation that the eGFP cassette had remained intact and was present in a single copy in both males. (e) Detection of eGFP and *Hprt* expression by qPCR compared to +/+ control. For both X/Y-Cas9 samples two different tissue samples were used, liver and kidney.

To determine whether the males encoded a functional Cas9, I set up two X/Y-Cas9 males in matings with hemizygous sgRNA/+ females (Figure 35a). There are four possible offspring genotypes: female mCherry+eGFP- (X/X;sgRNA/+), female mCherry-eGFP- (X/X;+/+), male mCherry+eGFP+ (X/Y-Cas9;sgRNA/+) male mCherry-eGFP+ (X/Y-Cas9;+/+). If co-inheritance of the Y-Cas9 transgene and the sgRNA transgene was embryonic lethal due to CRISPR-Cas9 induced mutations at *Top1*, then I would expect no male mCherry+eGFP+ offspring to be born.

The pups born from the matings were phenotyped for sex and assayed by *in vivo* imaging for mCherry expression at PN4. A total of 24 pups were born from three litters, and all four genotypes were present. Of the 24 pups born, nine (37%) were female sgRNA/+ heterozygous, and six (25%) were female mCherry negative (Figure 35b). All females were eGFP negative. Six males (25%) were X/Y-Cas9 and mCherry negative. However

three males (13%) were X/Y-Cas9;sgRNA/+ (Figure 35b). The occurring ratio of genotypes was not a statistically-significant deviation from Mendelian frequency. Overall this data suggests that the Y-linked Cas9-eGFP allele does not express Cas9, despite the constitutive CAG promoter and removal of the potentially inhibitory floxed-PGK-neomycin cassette.

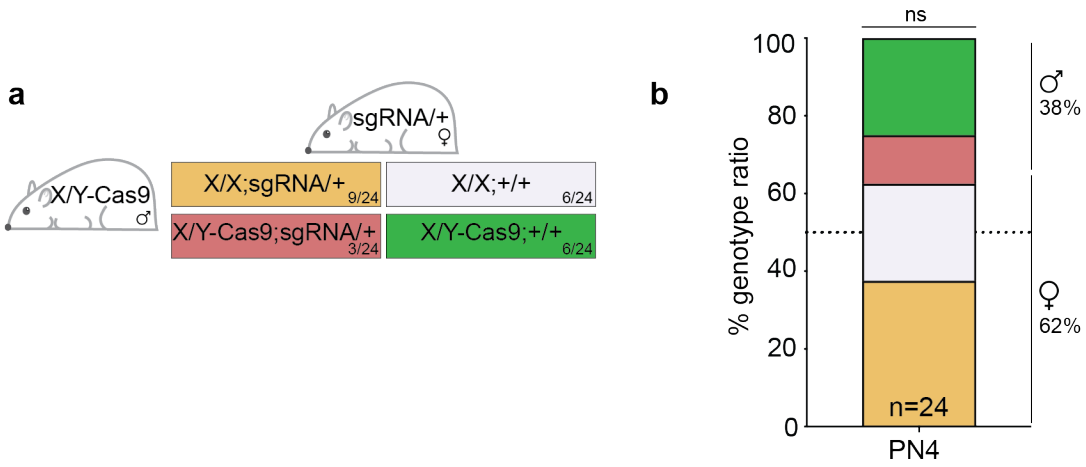

Figure 35. X/Y-Cas9 (-neo) matings to sgRNA/+ heterozygous females

(a) Mating strategy to assess if mCherry+eGFP+ offspring are embryonic lethal. X/Y-Cas9 males were set up for matings with sgRNA/+ heterozygous females. Four offspring genotypes were possible: female mCherry+eGFP- (X/X;sgRNA/+), female mCherry-eGFP- (X/X;+/+), male mCherry+eGFP+ (X/Y-Cas9;sgRNA/+) male mCherry-eGFP+ (X/Y-Cas9;+/+). (b) Offspring genotype ratios, assessed by phenotypic sex, *in vivo* imaging, and standard genotyping.

6.3 Discussion

The data presented in this chapter showed that transgenes can be successfully integrated into the X chromosome *Hprt* locus. In the mouse, targeting *Hprt* is extremely convenient as loss-of-function of *Hprt* has no detrimental phenotype (Kuehn et al., 1987, Hooper et al., 1987, Koller et al., 1989, Jinnah et al., 1990). I used a PGK-neomycin cassette for positive selection of transgenic clones, instead of 6-tg selection. Therefore, the neomycin was used as a target for copy number analysis and could be later excised by Cre recombination. Only two of the nine X-Cas9 clones appeared to have multiple or off-target integrations of the transgene. To determine if the multiple integrations were random integrations or transgene concatemers on the X chromosome, Southern blots could be performed. In ESCs and in a F1 generation X-Cas9/X female, the Cas9 transgene was expressed. In this X-Cas9/X female, the Cas9 transgene was hemizygous and therefore single-copy. Furthermore, the single-copy transgene may have been susceptible to X chromosome inactivation, silencing the transgene in some cells. Therefore, looking forward, I predict that male F2 generation X-Cas9 offspring may show a greater level of transgene expression, than the X-Cas9/X F1 mother.

The use of *Hprt* for knock-in of transgenes is extremely useful for generating female-specific mutations in the mouse, when the transgene, such as Cas9 or sgRNA, is carried by the father. Generating single-sex litters by sex-selective non-viability is immediately translatable for reducing mouse culling of the unrequired sex, in line with the Home Office 3Rs. Furthermore, the CRISPR-Cas9 bi-component system is translatable to other research groups that wish to use this system to induce sex-specific mutations at a gene of interest. However when translating the bi-component system to other species, it may be more challenging, if loss-of-function targeting of *Hprt* is not possible. In rabbits, heterozygous females are viable and fertile, however hemizygous knock-out males are embryonic lethal (Yin et al., 2015). Therefore, in the rabbit, male-carrying *Hprt* knock-in transgenes could not be generated. Conversely, loss-of-function *Hprt* in the male rat appears to have no detrimental phenotype on viability or fertility. However *Hprt* null male rats do show dysfunction of purine and nucleotide metabolism (Meek et al., 2016). In

cows, it is known that *Hprt* is expressed during pre-implantation development (Kita and Imai, 1993) however there have been no studies to date investigating the loss-of-function phenotype.

Although *Hprt* is an extremely useful docking-site in the mouse, direct translation of the technology to other species does not have to follow the exact-same strategy of knock-in. For example in cows, *Hprt* is expressed in the pre-implantation embryo (Kita and Imai, 1993). Therefore, the *Hprt* promoter could be utilised to drive transgene expression and there would be no loss-of-function of *Hprt.* Furthermore, transgenes could also be targeted to an intronic region of *Hprt,* thereby leaving the coding-sequence of *Hprt* intact. These strategies are not limited to *Hprt,* but other pre-implantation expressed gene promoters could be harnessed to drive transgene expression.

The data presented in this chapter showed that transgenes can be successfully integrated into the Y chromosome, at an intergenic region downstream of *Uty* and *Eif2s3y*. However, the efficiency of targeting at this location was extremely poor, with a single clone from 384 containing the targeted insertion. The poor efficiency of knock-in could be for two reasons. Firstly, the homology arms (HAs) in the Y-Cas9 TV were smaller than for the X-Cas9 targeting. The left HA was approximately 800bp, whilst the right HA was approximately 600bp. In the X-Cas9 targeting, both HAs were greater than 1kb. The reason for the smaller HAs in the Y-targeting was due to the highly repeat-rich nature of the *Uty-Eif2s3y* intergenic region. I generated HAs that not did contain highly repetitive regions, to reduce the potential risk of insert concatermerisation. The second reason is that although the HAs did not contain repeats, the intergenic region is highly complex and may be inhibitory to HDR. Therefore, irrespective of HA length, the rate of HDR may have been poor at this region. Nevertheless, the Y-chromosome targeting was successful, and injection of the Y-Cas9 ESC clone into blastocysts generated a stable transgenic mouse line.

The initial rationale for targeting downstream of Y-linked genes *Uty* and *Eif2s3y* was because these two genes are highly expressed in the developing embryo. Therefore, I predicted that successful integration of a transgene at this transcriptionally permissive

region of the Y chromosome would increase the likelihood of Cas9 expression. However no Cas9 expression was detected. One explanation for the lack of Cas9 expression from the Y-intergenic region, could be transgene silencing. The dsDNA plasmid vector used to generate the knock-in by HDR is bacterial DNA. Integration of bacterial DNA plasmid backbone at the target site, adjacent to the construct, may induce endogenous mechanisms to silence the entire transgene. Support for this theory comes in the form of boundary PCR genotyping. In the results of this chapter, I was able to successfully amplify the 5' boundary PCR product. However amplification of the 3' boundary PCR product consistently failed. This suggests that there could be some further backbone integration of the TV after the 3' HA. The reverse primer binding sites in the endogenous locus may have been lost, or are significantly further away from the transgene forward than predicted by *in silico* expected knock-in sequence. To investigate the sequence surrounding the integration site further, whole genome sequence or targeted high-throughput sequencing should be performed. If the sequencing results show that there is plasmid backbone present adjacent to the X-Cas9 cassette, it can be removed by CRISPR-Cas9 genome editing.

Anecdotal evidence also suggested that the presence of a PGK-neomycin cassette may induce local transgene silencing (personal communication, Lovell-Badge lab). To circumvent this risk, I used a constitutive X-Cre to excise the floxed-neomycin cassette. Prior to the neomycin excision, there was no transgene expression. Post-neomycin excision, there appeared to be eGFP expression by qPCR. However Cas9 expression was not able to fully induce a male-specific lethality effect *in vivo*, suggesting that the transgene expression is negligible.

Given the lack or negligible expression of the Y-Cas9 transgene in ESCs and in the mouse, future directions will be to repeat the Y-Cas9 targeting. One possibility for this repeat targeting is to the *Uty* locus. *Uty* is a constitutively expressed gene, both in ESCs and during embryo development. Loss-of-function of *Uty* is thought to be non-detrimental to the mouse (Shpargel et al., 2012). Previous studies have successfully targeted *Uty* for eGFP reporter tagging, in-frame with the ATG start codon of the *Uty* reading frame (Wang et al., 2013a). Targeting the *Uty* locus was also shown in the 2019

study by Yosef *et al* who generated a Y-linked sgRNA transgenic mouse line. In this study, Yosef *et al* targeted *Uty* introns to integrate the transgene (Yosef et al., 2019). Intronic or in-frame *Uty* targeting by HDR with the Cas9-eGFP transgene could keep *Uty* expression intact whilst allowing Cas9-eGFP expression.

Chapter 7. Summary

In conclusion, in this thesis I utilised and generated genetic tools to produce single-sex litters. Generating single-sex litters may have advantageous applications in agriculture, pest control, and in a laboratory research setting. In agriculture and in research, currently the unrequired sex is being culled needlessly after birth, generating a widespread animal welfare problem. I have described the potential of two different technologies to generate single-sex litters in order to reduce post-natal animal culling.

In Chapter 3, I described a method of generating single-sex litters using pre-existing mouse lines, as a proof-of-principle. The mouse lines encoded an autosome-linked inducible diphtheria toxin A, and an X-linked Cre recombinase. Inheritance of both the Cre recombinase, and the toxin resulted in female-specific lethality and all-male litters. Importantly, these results showed that it was possible to generate single-sex litters without impacting the viability of all the required-sex siblings.

In Chapters 4 and 5, I described the development of CRISPR-Cas9 tools, where the Cas9 and sgRNA components are genetically segregated. I showed that the bi-component system, targeting essential gene *Top1,* was sufficient to generate indel mutations *in vitro*. Furthermore, the mutations at *Top1* were stereotypic and generated frame-shift mutations, resulting in loss-of-function of *Top1*. The method of screening sgRNA mutation efficiency *in vitro* could be used for evaluating the efficiency of any sgRNA. The success of the *in vitro* bi-component system was recapitulated *in vivo.* Co-inheritance of the Cas9 and sgRNA transgenes induced mutations at *Top1* at a significantly higher rate to Cas9-only littermates. Importantly, the Cas9/sgRNA embryos were non-viable, although curiously, the litter size was largely unaffected. Therefore, I concluded that the CRISPR-Cas9 bi-component system was sufficient to introduce mutations at the sgRNA-target gene *in vitro* and *in vivo* and induce embryonic lethality. The CRISPR-Cas9 component is not limited to essential housekeeping genes, and could be used to target any gene of interest and therefore may be of interest to many research groups.

In Chapter 6, I generated sex chromosome-encoding Cas9 transgenic ESC lines, and used the targeted ESCs to generate transgenic mice. Generating X- and Y-linked Cas9 transgenes would allow for sex-specificity of the bi-component system. Generation of the X-Cas9 was successful, with Cas9 expression in ESCs. Furthermore, X-Cas9 chimeras successfully germline transmitted the X-Cas9 transgene to female offspring that also expressed Cas9. Unfortunately, although correctly targeted at the 5' end, the Y-linked Cas9 did not express either in ESCs or in mouse tissues. Nonetheless, this is the first example of generating X-linked Cas9-encoding lines, and the first example of generating autosome-linked sgRNA-expressing lines. These new transgenic mouse lines may be immediately useful for research laboratories.

The technology is not limited to targeting genes with essential housekeeping function, but any gene of interest in the genome could be targeted. The sgRNA-expressing mouse line was generated by targeting the transgene to the permissive locus *Hipp11* on mouse chromosome 11. The same targeting strategy could be used to generate a sgRNA transgene knock-in targeting any gene. This is highly translatable to other species, as the sgRNA transgene is autosome linked. As well as *H11* and *Rosa26* in the mouse, similar autosome-linked permissive-loci have been detected in other species, for example *Rosa26* is conserved in rats (Kobayashi et al., 2012), pigs (Kong et al., 2014, Li et al., 2014b, Li et al., 2014c), rabbits (Yang et al., 2016), and cows (Wang et al., 2018).

Looking forward, there are multiple other methods of generating single-sex litters by modifying the bi-component system. One possible future strategy to generate single-sex litters would be to produce X- or Y-linked sgRNA transgenes. In this scenario, the sgRNA would be uniquely inherited by daughters or sons, respectively. The sex chromosome-linked sgRNA mouse line would be crossed with the autosomal homozygous *R26*-Cas9 (Platt et al., 2014) mouse line. If the sgRNA is X-linked, there are two possible genotype outcomes: X-sgRNA/X;*R26*-Cas9/+ (female) or X/Y;*R26*-Cas9/+ (male). Only the female offspring inherit both transgenes, and therefore contain both CRISPR-Cas9 components (Table 6).

If the sgRNA is Y-linked, there are two possible genotype outcomes: X/X;*R26*-Cas9/+ (female), or X/Y-sgRNA;*R26*-Cas9/+ (male). In this scenario, only male offspring inherit both transgenes/CRISPR-Cas9 components (Table 7).

		Female	
		***R26*-Cas9**	***R26*-Cas9**
Male	**X-sgRNA**	X-sgRNA/X;*R26*-Cas9/+ (embryonic lethal)	X-sgRNA/X;*R26*-Cas9/+ (embryonic lethal)
	Y	X/Y;*R26*-Cas9/+	X/Y;*R26*-Cas9/+

Table 6. Hemizygous X-sgRNA/Y male mating to homozygous *R26*-Cas9 female

		Female	
		***R26*-Cas9**	***R26*-Cas9**
Male	**X**	X/X;*R26*-Cas9/+	X/X;*R26*-Cas9/+
	Y-sgRNA	X/Y-sgRNA;*R26*-Cas9/+ (embryonic lethal)	X/Y-sgRNA;*R26*-Cas9/+ (embryonic lethal)

Table 7. Hemizygous X/Y-sgRNA male mating to homozygous *R26*-Cas9 female

The sgRNA targeting *Top1* exon 15 has been shown in Chapters 4 and 5 to be highly efficient, generating loss-of-function mutations and embryonic non-viability. A construct expressing the *Top1* sgRNA and mCherry reporter can be used for targeting to the X- and Y chromosome. I would predict that co-inheritance of the X- or Y-linked sgRNA transgene with autosomal Cas9 would result in embryonic lethality.

Single-sex litters could also be generated by targeting genes necessary for male- or female-specific sex determination. In Chapter 2, I utilised a non-lethal sgRNA targeting the male TDF/sex-determination gene *Sry.* The same sgRNA could be used for generating sgRNA-transgene knock ins. The X-sgRNA(*Sry*) transgene would be biallelically carried

by the mother and therefore would be inherited by all offspring (Table 8). The X-linked sgRNA(*Sry)* transgenic females would be mated to *R26*-Cas9 homozygous males.

Conversely to earlier described strategies where inheritance of both the sgRNA and Cas9 transgenes is sex-specific, in this strategy, all offspring inherit both the sgRNA(*Sry)* and Cas9 transgenes. The *Sry* gene is uniquely carried by male offspring on the male-specific Y chromosome, therefore the *Sry* gene can only be targeted for knock-out in males (Table 8). Co-inheritance of the CRISPR-Cas9 components results in loss-of-function of *Sry* and male-to-female sex reversal. This strategy would give rise to all-female litters however approximately half of the litter would be genetically XY. The same result could also be produced by utilising autosomal sgRNA transgenes crossed with autosomal Cas9 transgenic lines, as only the Y-carrying males will be affected (Table 9). Although generating all-female litters by *Sry* knock-out may be a useful strategy for agriculture or pest control applications, researchers may be apprehensive to utilise this strategy in the laboratory. Although the litters will be all-female, the underlying genetic complement differs to the phenotypic sex, therefore the influence of the Y chromosome in phenotypically-female offspring could be a confounding factor in interpreting biological results.

In each of the strategies presented here, the Cas9/sgRNA offspring are transgenic and are embryonic lethal, or carry mutations in the target locus. The single-component littermates, e.g. Cas9/+ or sgRNA/+ would not have mutations but would still be transgenic.

		Female	
		X-sgRNA(*Sry)*	**X-sgRNA(*Sry)***
Male	***R26*-Cas9**	X-sgRNA(*Sry)*/X;*R26*-Cas9/+	X-sgRNA(*Sry)*/X;*R26*-Cas9/+
	***R26*-Cas9**	X-sgRNA(*Sry)*/Y;*R26*-Cas9/+ (*Sry* knock-out)	X-sgRNA(*Sry)*/Y;*R26*-Cas9/+ (*Sry* knock-out)

Table 8. X-sgRNA(*Sry*) homozygous female mating with *R26*-Cas9 homozygous males

		Female	
		***H11*-sgRNA(*Sry*)**	***H11*-sgRNA(*Sry*)**
Male	***R26*-Cas9**	X-sgRNA(*Sry*)/X;*R26*-Cas9/+	X-sgRNA(*Sry*)/X;*R26*-Cas9/+
	***R26*-Cas9**	X-sgRNA(*Sry*)/Y;*R26*-Cas9/+ (*Sry* knock-out)	X-sgRNA(*Sry*)/Y;*R26*-Cas9/+ (*Sry* knock-out)

Table 9. Autosomal H11-sgRNA(*Sry*) homozygous female mating with *R26*-Cas9 homozygous males

The fact that surviving animals are transgenic holds concerns for agricultural translation, regarding the safety of genetically modified animals and animal products. One method to circumvent the issue of transgenic animal products is by taking advantage of maternally deposited mRNAs in the early embryo. In this strategy, an autosomal Cas9, e.g. *R26*-Cas9 (Platt et al., 2014) would be carried by the mother, mono-allelically. Therefore, approximately half of the offspring inherit the Cas9 transgene. The other half of the offspring are not transgenic for Cas9; however all zygotes are pre-loaded with Cas9 mRNA from the oocyte. Previous studies have shown that zygotic maternal-Cas9 mRNA generates higher rates of genome edits at target loci, compared to when Cas9 is supplied by mRNA or protein microinjection (Cebrian-Serrano et al., 2017).

The mono-allelic Cas9/+ mother would be crossed with a sex chromosome-linked sgRNA expressing male. If the sgRNA is X-linked, the females inherit the sgRNA transgene (Table 10). If the sgRNA is Y-linked, the males inherit the sgRNA transgene (Table 11). In each case, there are four offspring genotypes. Inheritance of both the sgRNA and Cas9 components, either genetically or by maternally deposited transcripts, results in mutations in the target locus, inducing non-viability (or fertility). One-quarter of the litter/half of the surviving population is genetically wildtype.

		Female	
		***R26*-Cas9**	+
Male	**X-sgRNA**	X-sgRNA/X;*R26*-Cas9/+ (embryonic lethal)	X-sgRNA/X;+/+ (maternally loaded Cas9 mRNA, embryonic lethal)
	Y	X/Y;*R26*-Cas9/+	X/Y;+/+ (wildtype)

Table 10. X-linked sgRNA male mating to hemizygous *R26*-Cas9 to generate wildtype offspring

		Female	
		***R26*-Cas9**	+
Male	**X**	X/X;*R26*-Cas9/+	X/X;+/+ (wildtype)
	Y-sgRNA	X/Y-sgRNA;*R26*-Cas9/+ (embryonic lethal)	X/Y-sgRNA;+/+ (maternally loaded Cas9 mRNA, embryonic lethal)

Table 11. Y-linked sgRNA mating with heterozygous *R26*-Cas9 to generate wildtype offspring

There is still public and political discussion as to the safety and efficacy of genetically modified animals or animal products for human consumption. The main advantage of the CRISPR-Cas9 system described in this thesis, is that the X- or Y-linked Cas9 endonuclease gene is lost with the embryonic-lethal population. The surviving population carries only the sgRNA transgene, which is not protein-coding. The fact that the surviving population carry a genetic modification which is not protein-coding, may provide an acceptable alternative for consumption of genetically modified animals. I have also described in this thesis alternative methods for inducing sex-specific lethality, whilst retaining half of the surviving litter completely wildtype, by utilising maternally deposited mRNAs. The feasibility of the wildtype-generating bi-component system warrants further investigation for my long-term future experiments. Furthermore, my long-term future experiments also include the generation of X-linked sgRNA transgenic

animals, including the *Top1* sgRNA transgene, and the *Sry* sgRNA transgene. These two mouse models can be utilised as earlier described.

In the short-term future I believe it will also be important to follow up on the Y-Cas9 experimental matings, using the sgRNA mouse line. It will be important to investigate further the number of live births of males that are genotype Y-Cas9/sgRNA, to determine if co-inheritance of these two alleles results in non-viability of males and female-bias sex skew. Furthermore, if it is shown that the Y-Cas9 is functional to reduce the number of males born, thereby skewing offspring ratios, exactly when in development the Y-Cas9 transgene is functional. Y-Cas9 transgene expression timing can be investigated using immunofluorescence techniques. Next, it will be essential to investigate the functionality of the X-Cas9 transgene, by setting up mouse matings between X-Cas9 hemizygous males, and homozygous sgRNA females. If the X-Cas9 transgene is functional, I predict that these matings will produce all-male litters. It may be necessary to determine that the X-linked Cas9 transgene is expressed during multiple stages of embryonic development by molecular biology techniques. Lastly, it may be required to investigate the position of the *H11 Top1* sgRNA transgene, for example by performing CRISPR-Cas9 enhanced Nanopore sequencing.

In summary, I have shown in this thesis that it is possible to utilise a bi-component system to generate single-sex litters, and have made significant progress towards generating the genetic tools for a CRISPR-Cas9 bi-component system. Utilising a CRISPR-Cas9 bi-component system carries strengths in that the mouse models can be simply bred and maintained as standard breedings. Therefore, the models generated here may be immediately applicable to laboratory use for generating single-sex litters, in line with the Home Office 3Rs, if the sex chromosome linked Cas9 transgenes are shown to be functional. In this laboratory application the CRISPR-Cas9 bi-component strategy strengths is that the Cas9 and sgRNA lines can be maintained and bred separately, to ensure transmission of the transgenes, or in experimental matings to produce single-sex litters. Furthermore, different sgRNA transgenic mouse lines may be generated to target any gene of interest, thereby becoming applicable to many research groups to knock-out genes in a sex-specific manner. However for the agricultural approach, transgenic

breeders carrying a Cas9 or sgRNA transgene would first have to be generated, thereby creating a challenge for immediate translation of the technology. However once generated, it may be possible to sell Cas9-transgenic sperm to livestock breeders, to generate sex-selected offspring by *in vitro* fertilisation with sgRNA-carrying oocytes. In this strategy, farmers can maintain sgRNA and wildtype females, and buy Cas9-carrying sperm as required. However the main weakness of the agricultural translation is the use of GM animals to generate animal produce, which may be unpopular with the consumer. More realistically however, the CRISPR-Cas9 bi-component system may be more relevant to pest control. Many invasive species are rodents, and therefore genetically modified bi-component rodents may be released into the natural population at high frequency, to produce all-female or all-male litters.

Chapter 8. Appendices

8.1 Appendix A: Media and buffers

	Concentration (mM)	
Components	**Follicle holding medium (FHM)**	**Potassium simplex optimised medium (KSOM)**
NaCl	95	95
KH_2PO_4	2.5	2.5
$MgSO_4$	0.2	0.2
Lactate	10	10
Pyruvate	0.2	0.2
Glucose	0.2	0.2
Glutamine	1	1
Bovine serum albumin	1mg/ml	1mg/ml
EDTA	0.01	0.01
$NaHCO_3$	4	25
HEPES	20	-
$CaCl_2$	1.71	1.71

Table 12. Embryo media

Medium	Components	Volume	Final conc	Product code
2i + LIF	NDiff227	489ml	-	Cellartis/Takara-Clontech (Y40002)
	Penicillin/Streptomycin	5ml	50U/ml, 50µg/ml	Thermo Fisher Scientific (150700A973163)
	Glutamax	5ml	2mM	Thermo Fisher Scientific (35050038)
	Human recombinant LIF	50µl	1000U/ml	Merck-Millipore (ESG1107)
	CHIR99021	160µl	1uM	Axon Medchem (1386)
	PD0325901	160µl	3uM	Axon Medchem (1408)
	Human recombinant serum albumin	500µl	50µg/ml	Sigma-Aldrich (A9731)
	2-mercaptoethanol	3.2µl	0.1uM	Sigma-Aldrich (M6250)
Freezing medium	Dimethyl (DMSO)	1ml	20%	
	2i+LIF	4ml	80%	
FACS medium	Fetal calf serum		2%	
	2i+LIF		98%	

Table 13. Cell culture media

Medium	Components	Concentration	Volume	Final conc
Bradley lysis buffer	Tris-HCl (pH7.5)	1M	5ml	10mM
	EDTA	0.5M	10ml	10mM
	SDS	10%	25ml	0.5%
	NaCl	5M	1ml	10mM
	H_2O	-	459ml	-
+Proteinase K	ProteinaseK	20mg/ml		1mg/ml
10X KT buffer	Tris-HCl (pH9.1)	670mM		
	$(NH_4)_2SO_4$	160mM		
	$MgCl_2$	35mM		
	BSA	1.5mg/ml		
KT lysis buffer	10X KT buffer		10µl	1X
	NP40	10%	5µl	1%
	ProteinaseK	20mg/ml	4µl	1mg/ml
	H_2O		81µl	

Table 14. DNA extraction buffers

S.O.C. medium	**Concentration**	**Total for 10 litres**
Part A (autoclave sterilise)		
Bacto Tryptone	20g	200g
NaCl	0.584g	5.840g
KCl	0.186g	1.860g
Water	0.800 l	8 l
Part B (filter sterilise)		
$MgCl_2.6H_2O$	2.033g	20.330g
$MgSO_4.7H_2O$	2.464g	24.640g
Glucose	3.603g	36.030g
Water	0.2 l	2l
Final product		
Part A	8ml	
Part B	2ml	

Table 15. S.O.C. medium

Medium	Components	Concentration
Radioimmunoprecipitation assay (RIPA) buffer	Tris-Cl (pH8)	10mM
	EDTA	1mM
	EGTA	0.5mM
	Triton-X 100	1%
	Sodium deoxycholate	0.1%
	Sodium dodecyl sulphate (SDS)	0.1%
	NaCl	140mM
	Phenylmethane sulphonyl fluoride (PMSF)	1mM
Protein extraction buffer	RIPA buffer	10ml
	Phosphatase inhibitor	1x tablet
	Protease inhibitor	1x tablet
	PMSF	100µl
Laemmli buffer	Tris-Cl (pH6.8)	60mM
	SDS	2%
	Glycerol	10%
	2-mercaptoethanol	5%
	Bromophenol blue	0.01%
Running buffer	Tris base	3g/L
	Glycine	14.4g/L
	SDS	1g/L
Transfer buffer	Tris base	3g/L
	Glycine	14.4g/L
	SDS	50mg/L
TBS-Tween	Tris-Cl (pH7.5)	20mM
	NaCl	150mM
	Tween-20	0.2%

Table 16. Protein biology buffers

Medium	Components	Final conc	Product code
Depurination buffer	HCl	0.25M	
	Water		
Denaturation buffer	NaCl	1.5M	
	NaOH	0.5M	
	Water		
Neutralisation buffer (pH 7.5)	NaCl	1.5M	
	Tris	0.5M	
Hybridisation buffer			#11603558001 Roche DIG Easy Hyb
Washing buffer	Maleic acid buffer	1x	
	Tween-20	0.3%	
Maleic acid buffer	Maleic acid	0.1M	M0375 Sigma
	NaCl	0.15M	
	NaOH pellets	To pH 7.5	
10 X Blocking buffer	Blocking powder	10% (w/v)	#11363514910 detection kit Roche
Detection buffer	Tris-HCl	0.1M	
	NaCl	0.1M	

Table 17. Southern blot buffers

8.2 Appendix B: Oligonucleotides

Name	Sequence	Reference
Sly_F	CAGTTACCAATCAACACATCAC	
Sly_R	CTGGAGCTCTACAGTGATGA	
Myog_F	TTACGTCCATCGTGGACAGCAT	
Myog_R	TGGGCTGGGTGTTAGTCTTAT	
Hprt_Cre_F	TTCATAGAGACAAGGAATGTGTCC	
Hprt_Cre_MutR	CTCGTGCTTTACGGTATCGC	
Hprt_Cre_wtR	AATCCAGCAGGTCAGCAAAG	
*R26*_DTA_F	AAAGTCGCTCTGAGTTGTTAT	(Soriano, 1999) (Srinivas et al., 2001)
*R26*_DTA_MutR	GCGAAGAGTTTGTCCTCAACC	
*R26*_DTA_wtR	GGAGCGGGAGAAATGGATATG	
Cas9_RT_F	AAACAGCAGATTCGCCTGGA	(Ran et al., 2013b)
Cas9_RT_R	TCATCCGCTCGATGAAGCTC	
eGFP_RT_F	CTACCCCGACCACATGAAG	(Yoshimi et al., 2016)
eGFP_RT_R	CTTGTGCCCCAGGATGTT	
mCherry_F	ATTACCGGTCGCCACCATGGTGAGC AAG	
mCherry_R	TATGAATTCTCAGGACTTGTACAGCT CGTCCATG	
mCherry-mut_R	TATCCAGCCCATGGTTTTCTTCTGCA TTACGGG	
Sry_sgRNA	CCACGCATTTATGGTGTGGTCCCG	Turner lab (unpublished)
Sry_sgRNA_RC	AAACCGGGACCACACCATAAATGC	
Top1_sgRNA1	CCACGGGGCTGCTGTTCACTTAGAG	
Top1_sgRNA1_RC	AAACCTCTAAGTGAACAGCAGCCCC	
Top1_sgRNA2	CCACGCGATCAAAAAGATCGTCCTC	
Top1_sgRNA2_RC	AAACGAGGACGATCTTTTTGATCGC	

Top1_sgRNA3	CCACGAACACAAAGATCGAGAACAC	
Top1_sgRNA3_RC	AAACGTGTTCTCGATCTTTGTGTTC	
Top1_g1_F	GAAGGAGAGACGGCAGACAC	
Top1_g1_R	TGCAGAACATGCAAAAGCCC	
Top1_g1_MiSeqF	TCGTCGGCAGCGTCAGATGTGTATA AGAGACAGGAAGGAGAGACGGCAG ACAC	
Top1_g1_MiSeqR	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGTGCAGAACATGCAAAAG CCC	
Top1_g2_F	ACCACAAATGGCTGAGAACTGA	
Top1_g2_R	GGTCTGCTGCTGGTTACAGA	
Top1_g2_MiSeqF	TCGTCGGCAGCGTCAGATGTGTATA AGAGACAGACCACAAATGGCTGAGA ACTGA	
Top1_g2_MiSeqR	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGGGTCTGCTGCTGGTTAC AGA	
Top1_g3_F	TTGAGGCAAGGCAATGGGAT	
Top1_g3_R	ACTTTTCCCGGTCCTTATCCTT	
Top1_g3_MiSeqF	TCGTCGGCAGCGTCAGATGTGTATA AGAGACAG TTGAGGCAAGGCAATGGGAT	
Top1_g3_MiSeqR	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAG ACTTTTCCCGGTCCTTATCCTT	
attB_F1	ATTGGCGCGCCGGTTCCTGGCCTTTT GCT	
attB_R1	TATACTAGGGGAGTGGCCAACTCCA TCA	
attB_F2	ATTGGCGCGCCCATTCTCGAGCATGC GCCATTC	

attB_R2	TATACTAGTCCTGCGGCCGCTCCC	
attB_F3	ATTACTAGTGCTCGAGAATGGCGCA TGTGAGGGC	
attB_R3	GAAATAGGCCCTCTCTAGAGCCATTT GTC	
Hprt_MiSeqF	TCGTCGGCAGCGTCAGATGTGTATA AGAGACAGGCAGATTAGCGATGATG AACC	
Hprt_MiSeqR	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGC CAGCAAGAGACACTGATTCAAGG	
X-Cas9_5'F	GAAACCTGGGTGTGATAGGCTT	
X-Cas9_5'R	AGGTCATGTACTGGGCACAA	
X-Cas9_3'F	GACAACCAGGAATAGCCAGTACATC	
X-Cas9_3'R	CAACACACCAGCTCAACCAA	
Sry_MiSeqF	TCGTCGGCAGCGTCAGATGTGTATA AGAGACAG TCTGAAGAAGAGACAAGTTTTGGG	
Sry_MiSeqR	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAG GTGACACTTTAGCCCTCCGAT	
Y_Chr_MiSeqF	TCGTCGGCAGCGTCAGATGTGTATA AGAGACAGGTGAAGGCTGCCCATGA ATTCAA	
Y_Chr_MiSeqR	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGTAAGTCATCTCTGCATG TGTCGC	
Y-Cas9_F	GACAACCAGGAATAGCCAGTACATC	
Y-Cas9_R	AGGTCATGTACTGGGCACAA	
Y-Cas9 minusNeo_F	CTACCCCGACCACATGAAG	
Y-Cas9 minusNeo_R	CAGATCAATTGAGGAAACTGACAAG C	

Table 18. Oligonucleotides

8.3 Appendix C: TaqMan probes

Gene (gene expression assay)	**Product code (all by Thermo Fisher Scientific)**
Hprt	Mm03024075_m1
Eif2s3y	Mm01210630_m1
Uty	Mm00447710_m1
Gapdh	Mm99999915_g1
Pou5f1	Mm03053917_g1
eGFP	Mr04097229_mr
mCherry	Mr07319438_mr
Gene (copy number assay)	
Tfrc	Copy number reference gene 4458370
Hprt	Mm00522878_cn
eGFP	Mr00660654_cn
Neomycin	Mr00299300_cn

Table 19. TaqMan probes

8.4 Appendix D: Antibodies

Protein	**Product code**
CAS9	Novus Bio NBP2-36440
TOP1 (N-terminal)	Abcam Ab109374
TOP1 (C-terminal)	Abcam Ab245432
TUBULIN	Sigma T9026
GAPDH	Santa Cruz Biotechnology Sc-25778

Table 20. Antibodies

Reference List

ADLER, D. A., RUGARLI, E. I., LINGENFELTER, P. A., TSUCHIYA, K., POSLINSKI, D., LIGGITT, H. D., CHAPMAN, V. M., ELLIOTT, R. W., BALLABIO, A. & DISTECHE, C. M. 1997. Evidence of evolutionary up-regulation of the single active X chromosome in mammals based on Clc4 expression levels in Mus spretus and Mus musculus. *Proc Natl Acad Sci U S A,* 94**,** 9244-8.

ADRA, C. N., ELLIS, N. A. & MCBURNEY, M. W. 1988. The family of mouse phosphoglycerate kinase genes and pseudogenes. *Somat Cell Mol Genet,* 14**,** 69-81.

AITKEN, R. J. & MARSHALL GRAVES, J. A. 2002. The future of sex. *Nature,* 415**,** 963.

ALBRECHT, K. H. & EICHER, E. M. 2001. Evidence that Sry is expressed in pre-Sertoli cells and Sertoli and granulosa cells have a common precursor. *Dev Biol,* 240**,** 92-107.

ALLEN, F., CREPALDI, L., ALSINET, C., STRONG, A. J., KLESHCHEVNIKOV, V., DE ANGELI, P., PALENIKOVA, P., KHODAK, A., KISELEV, V., KOSICKI, M., BASSETT, A. R., HARDING, H., GALANTY, Y., MUNOZ-MARTINEZ, F., METZAKOPIAN, E., JACKSON, S. P. & PARTS, L. 2018. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol*.

AUGUI, S., FILION, G. J., HUART, S., NORA, E., GUGGIARI, M., MARESCA, M., STEWART, A. F. & HEARD, E. 2007a. Sensing X Chromosome Pairs Before X Inactivation via a Novel X-Pairing Region of the <em>Xic</em&gt. *Science,* 318**,** 1632.

AUGUI, S., FILION, G. J., HUART, S., NORA, E., GUGGIARI, M., MARESCA, M., STEWART, A. F. & HEARD, E. 2007b. Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. *Science,* 318**,** 1632-6.

BACHTROG, D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet,* 14**,** 113-24.

BACHTROG, D. 2014. Signs of genomic battles in mouse sex chromosomes. *Cell,* 159**,** 716-8.

BAE, S., KWEON, J., KIM, H. S. & KIM, J. S. 2014. Microhomology-based choice of Cas9 nuclease target sites. *Nat Methods,* 11**,** 705-6.

BAILEY, J. A., CARREL, L., CHAKRAVARTI, A. & EICHLER, E. E. 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc Natl Acad Sci U S A,* 97**,** 6634-9.

BARAKAT, T. S., GUNHANLAR, N., PARDO, C. G., ACHAME, E. M., GHAZVINI, M., BOERS, R., KENTER, A., RENTMEESTER, E., GROOTEGOED, J. A. & GRIBNAU, J. 2011. RNF12 activates Xist and is essential for X chromosome inactivation. *PLoS Genet,* 7**,** e1002001.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* 315**,** 1709-12.

BARRIONUEVO, F., BAGHERI-FAM, S., KLATTIG, J., KIST, R., TAKETO, M. M., ENGLERT, C. & SCHERER, G. 2006. Homozygous inactivation of Sox9 causes complete XY sex reversal in mice. *Biol Reprod,* 74**,** 195-201.

BAUER, H., WILLERT, J., KOSCHORZ, B. & HERRMANN, B. G. 2005. The t complex-encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nat Genet,* 37**,** 969-73.

BELFORT, M. & ROBERTS, R. J. 1997. Homing endonucleases: keeping the house in order. *Nucleic Acids Res,* 25**,** 3379-88.

BELL, C. E. & EISENBERG, D. 1996. Crystal Structure of Diphtheria Toxin Bound to Nicotinamide Adenine Dinucleotide. *Biochemistry,* 35**,** 1137-1149.

BELLOTT, D. W., HUGHES, J. F., SKALETSKY, H., BROWN, L. G., PYNTIKOVA, T., CHO, T. J., KOUTSEVA, N., ZAGHLUL, S., GRAVES, T., ROCK, S., KREMITZKI, C., FULTON, R. S., DUGAN, S., DING, Y., MORTON, D., KHAN, Z., LEWIS, L., BUHAY, C., WANG, Q., WATT, J., HOLDER, M., LEE, S., NAZARETH, L., ALFOLDI, J., ROZEN, S., MUZNY, D. M., WARREN, W. C., GIBBS, R. A., WILSON, R. K. & PAGE, D. C. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature,* 508**,** 494-9.

BELLOTT, D. W., SKALETSKY, H., CHO, T. J., BROWN, L., LOCKE, D., CHEN, N., GALKINA, S., PYNTIKOVA, T., KOUTSEVA, N., GRAVES, T., KREMITZKI, C., WARREN, W. C., CLARK, A. G., GAGINSKAYA, E., WILSON, R. K. & PAGE, D. C. 2017. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nat Genet,* 49**,** 387-394.

BELLOTT, D. W., SKALETSKY, H., PYNTIKOVA, T., MARDIS, E. R., GRAVES, T., KREMITZKI, C., BROWN, L. G., ROZEN, S., WARREN, W. C., WILSON, R. K. & PAGE, D. C. 2010. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature,* 466**,** 612-6.

BERLETCH, J. B., MA, W., YANG, F., SHENDURE, J., NOBLE, W. S., DISTECHE, C. M. & DENG, X. 2015. Escape from X inactivation varies in mouse tissues. *PLoS Genet,* 11**,** e1005079.

BERLETCH, J. B., YANG, F. & DISTECHE, C. M. 2010. Escape from X inactivation in mice and humans. *Genome Biol,* 11**,** 213.

BERLETCH, J. B., YANG, F., XU, J., CARREL, L. & DISTECHE, C. M. 2011. Genes that escape from X inactivation. *Human genetics,* 130**,** 237-245.

BERMEJO-ALVAREZ, P., PERICUESTA, E., MIRANDA, A., DE FRUTOS, C., PEREZ-CEREZALES, S., LUCIO, A., RIZOS, D. & GUTIERREZ-ADAN, A. 2011. New challenges in the analysis of gene transcription in bovine blastocysts. *Reprod Domest Anim,* 46 Suppl 3**,** 2-10.

BERNSTEIN, R. M. & MUKHERJEE, B. B. 1972. Control of nuclear RNA synthesis in 2-cell and 4-cell mouse embryos. *Nature,* 238**,** 457-9.

BIRCHLER, J. A. 2012. Claims and counterclaims of X-chromosome compensation. *Nat Struct Mol Biol,* 19**,** 3-5.

BISHOP, C. E., WHITWORTH, D. J., QIN, Y., AGOULNIK, A. I., AGOULNIK, I. U., HARRISON, W. R., BEHRINGER, R. R. & OVERBEEK, P. A. 2000. A transgenic insertion upstream of sox9 is associated with dominant XX sex reversal in the mouse. *Nat Genet,* 26**,** 490-4.

BOLOTIN, A., QUINQUIS, B., SOROKIN, A. & EHRLICH, S. D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology,* 151**,** 2551-61.

BOROVIAK, K., DOE, B., BANERJEE, R., YANG, F. & BRADLEY, A. 2016. Chromosome engineering in zygotes with CRISPR/Cas9. *Genesis,* 54**,** 78-85.

BOROVIAK, K., FU, B., YANG, F., DOE, B. & BRADLEY, A. 2017. Revealing hidden complexities of genomic rearrangements generated with Cas9. *Sci Rep,* 7**,** 12867.

BORSANI, G., TONLORENZI, R., SIMMLER, M. C., DANDOLO, L., ARNAUD, D., CAPRA, V., GROMPE, M., PIZZUTI, A., MUZNY, D., LAWRENCE, C., WILLARD, H. F., AVNER, P. & BALLABIO, A. 1991. Characterization of a murine gene expressed from the inactive X chromosome. *Nature,* 351**,** 325-9.

BROCKDORFF, N., ASHWORTH, A., KAY, G. F., MCCABE, V. M., NORRIS, D. P., COOPER, P. J., SWIFT, S. & RASTAN, S. 1992. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell,* 71**,** 515-26.

BROCKDORFF, N., KAY, G., SMITH, S., KEER, J. T., HAMVAS, R. M., BROWN, S. D. & RASTAN, S. 1991. High-density molecular map of the central span of the mouse X chromosome. *Genomics,* 10**,** 17-22.

BROWN, C. J., BALLABIO, A., RUPERT, J. L., LAFRENIERE, R. G., GROMPE, M., TONLORENZI, R. & WILLARD, H. F. 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature,* 349**,** 38-44.

BROWN, C. J., HENDRICH, B. D., RUPERT, J. L., LAFRENIERE, R. G., XING, Y., LAWRENCE, J. & WILLARD, H. F. 1992. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell,* 71**,** 527-42.

BURGOYNE, P. S. 1982. Genetic homology and crossing over in the X and Y chromosomes of Mammals. *Hum Genet,* 61**,** 85-90.

BURT, A. 2003. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc Biol Sci,* 270**,** 921-8.

CALLEBAUT, I., MALIVERT, L., FISCHER, A., MORNON, J. P., REVY, P. & DE VILLARTAY, J. P. 2006. Cernunnos interacts with the XRCC4 x DNA-ligase IV complex and is homologous to the yeast nonhomologous end-joining factor Nej1. *J Biol Chem,* 281**,** 13857-60.

CARREL, L. & WILLARD, H. F. 1999. Heterogeneous gene expression from the inactive X chromosome: an X-linked gene that escapes X inactivation in some human cell lines but is inactivated in others. *Proc Natl Acad Sci U S A,* 96**,** 7364-9.

CARREL, L. & WILLARD, H. F. 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature,* 434**,** 400-4.

CARY, R. B., CHEN, F., SHEN, Z. & CHEN, D. J. 1998. A central region of Ku80 mediates interaction with Ku70 in vivo. *Nucleic Acids Res,* 26**,** 974-9.

CATT, S. L., CATT, J. W., GOMEZ, M. C., MAXWELL, W. M. & EVANS, G. 1996. Birth of a male lamb derived from an in vitro matured oocyte fertilised by intracytoplasmic injection of a single presumptive male sperm. *Vet Rec,* 139**,** 494-5.

CEBRIAN-SERRANO, A., ZHA, S., HANSSEN, L., BIGGS, D., PREECE, C. & DAVIES, B. 2017. Maternal Supply of Cas9 to Zygotes Facilitates the Efficient Generation of Site-Specific Mutant Mouse Models. *PLOS ONE,* 12**,** e0169887.

CHABOISSIER, M. C., KOBAYASHI, A., VIDAL, V. I., LUTZKENDORF, S., VAN DE KANT, H. J., WEGNER, M., DE ROOIJ, D. G., BEHRINGER, R. R. & SCHEDL, A. 2004. Functional analysis of Sox8 and Sox9 during sex determination in the mouse. *Development,* 131**,** 1891-901.

CHAKRABARTI, A. M., HENSER-BROWNHILL, T., MONSERRAT, J., POETSCH, A. R., LUSCOMBE, N. M. & SCAFFIDI, P. 2019. Target-Specific Precision of CRISPR-Mediated Genome Editing. *Mol Cell,* 73**,** 699-713 e6.

CHAMBERS, I., COLBY, D., ROBERTSON, M., NICHOLS, J., LEE, S., TWEEDIE, S. & SMITH, A. 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell,* 113**,** 643-55.

CHAMPER, J., REEVES, R., OH, S. Y., LIU, C., LIU, J., CLARK, A. G. & MESSER, P. W. 2017. Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genet,* 13**,** e1006796.

CHAN, Y. S., NAUJOKS, D. A., HUEN, D. S. & RUSSELL, S. 2011. Insect population control by homing endonuclease-based gene drive: an evaluation in Drosophila melanogaster. *Genetics,* 188**,** 33-44.

CHANG, H. H., WATANABE, G. & LIEBER, M. R. 2015. Unifying the DNA end-processing roles of the artemis nuclease: Ku-dependent artemis resection at blunt DNA ends. *J Biol Chem,* 290**,** 24036-50.

CHARLESWORTH, B. & CHARLESWORTH, D. 2000. The degeneration of Y chromosomes. *Philos Trans R Soc Lond B Biol Sci,* 355**,** 1563-72.

CHAZAUD, C. & YAMANAKA, Y. 2016. Lineage specification in the mouse preimplantation embryo. *Development,* 143**,** 1063-74.

CHAZAUD, C., YAMANAKA, Y., PAWSON, T. & ROSSANT, J. 2006. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev Cell,* 10**,** 615-24.

CHEN-TSAI, R. Y. 2019. Using TARGATT Technology to Generate Site-Specific Transgenic Mice. *Methods Mol Biol,* 1874**,** 71-86.

CHIOU, S. H., WINTERS, I. P., WANG, J., NARANJO, S., DUDGEON, C., TAMBURINI, F. B., BRADY, J. J., YANG, D., GRUNER, B. M., CHUANG, C. H., CASWELL, D. R., ZENG, H., CHU, P., KIM, G. E., CARPIZO, D. R., KIM, S. K. & WINSLOW, M. M. 2015. Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing. *Genes Dev,* 29**,** 1576-85.

CHOI, J., CLEMENT, K., HUEBNER, A. J., WEBSTER, J., ROSE, C. M., BRUMBAUGH, J., WALSH, R. M., LEE, S., SAVOL, A., ETCHEGARAY, J. P., GU, H., BOYLE, P., ELLING, U., MOSTOSLAVSKY, R., SADREYEV, R., PARK, P. J., GYGI, S. P., MEISSNER, A. & HOCHEDLINGER, K. 2017. DUSP9 Modulates DNA Hypomethylation in Female Mouse Pluripotent Stem Cells. *Cell Stem Cell,* 20**,** 706-719 e7.

CHOWDHURY, M. M. R., LIANGUANG, X., KONG, R., PARK, B. Y., MESALAM, A., JOO, M. D., AFRIN, F., JIN, J. I., LIM, H. T. & KONG, I. K. 2019. In vitro production of sex preselected cattle embryos using a monoclonal antibody raised against bull sperm epitopes. *Anim Reprod Sci,* 205**,** 156-164.

CHU, V. T., WEBER, T., WEFERS, B., WURST, W., SANDER, S., RAJEWSKY, K. & KUHN, R. 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol,* 33**,** 543-8.

CLEGG, K. B. & PIKO, L. 1983. Poly(A) length, cytoplasmic adenylation and synthesis of poly(A)+ RNA in early mouse embryos. *Dev Biol,* 95**,** 331-41.

CLEMSON, C. M., MCNEIL, J. A., WILLARD, H. F. & LAWRENCE, J. B. 1996. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol,* 132**,** 259-75.

CLINTON, M., NANDI, S., ZHAO, D., OLSON, S., PETERSON, P., BURDON, T. & MCBRIDE, D. 2016. Real-Time Sexing of Chicken Embryos and Compatibility with in ovo Protocols. *Sex Dev,* 10**,** 210-216.

COCQUET, J., ELLIS, P. J., MAHADEVAIAH, S. K., AFFARA, N. A., VAIMAN, D. & BURGOYNE, P. S. 2012. A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. *PLoS Genet,* 8**,** e1002900.

COCQUET, J., ELLIS, P. J., YAMAUCHI, Y., MAHADEVAIAH, S. K., AFFARA, N. A., WARD, M. A. & BURGOYNE, P. S. 2009. The multicopy gene Sly represses the sex chromosomes in the male mouse germline after meiosis. *PLoS Biol,* 7**,** e1000244.

CONCORDET, J. P. & HAEUSSLER, M. 2018. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res,* 46**,** W242-W245.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

CONWAY, S. J., MAHADEVAIAH, S. K., DARLING, S. M., CAPEL, B., RATTIGAN, A. M. & BURGOYNE, P. S. 1994. Y353/B: a candidate multiple-copy spermiogenesis gene on the mouse Y chromosome. *Mamm Genome,* 5**,** 203-10.

COOKE, H. J. & SMITH, B. A. 1986. Variability at the telomeres of the human X/Y pseudoautosomal region. *Cold Spring Harb Symp Quant Biol,* 51 Pt 1**,** 213-9.

CORTEZ, D., MARIN, R., TOLEDO-FLORES, D., FROIDEVAUX, L., LIECHTI, A., WATERS, P. D., GRUTZNER, F. & KAESSMANN, H. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature,* 508**,** 488-93.

COSTANTINI, S., WOODBINE, L., ANDREOLI, L., JEGGO, P. A. & VINDIGNI, A. 2007. Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair (Amst),* 6**,** 712-22.

COTTON, A. M., PRICE, E. M., JONES, M. J., BALATON, B. P., KOBOR, M. S. & BROWN, C. J. 2015. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Hum Mol Genet,* 24**,** 1528-39.

CURTIS, C. F. 1968. Possible use of translocations to fix desirable genes in insect pest populations. *Nature,* 218**,** 368-9.

DAVIS, A. J. & CHEN, D. J. 2013. DNA double strand break repair via non-homologous end-joining. *Transl Cancer Res,* 2**,** 130-143.

DE LEON, V., JOHNSON, A. & BACHVAROVA, R. 1983. Half-lives and relative amounts of stored and polysomal ribosomes and poly(A) + RNA in mouse oocytes. *Dev Biol,* 98**,** 400-8.

DEJARNETTE, J. M., LEACH, M. A., NEBEL, R. L., MARSHALL, C. E., MCCLEARY, C. R. & MORENO, J. F. 2011. Effects of sex-sorting and sperm dosage on conception rates of Holstein heifers: is comparable fertility of sex-sorted and conventional semen plausible? *J Dairy Sci,* 94**,** 3477-83.

DELTCHEVA, E., CHYLINSKI, K., SHARMA, C. M., GONZALES, K., CHAO, Y., PIRZADA, Z. A., ECKERT, M. R., VOGEL, J. & CHARPENTIER, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature,* 471**,** 602-7.

DENG, X., HIATT, J. B., NGUYEN, D. K., ERCAN, S., STURGILL, D., HILLIER, L. W., SCHLESINGER, F., DAVIS, C. A., REINKE, V. J., GINGERAS, T. R., SHENDURE, J., WATERSTON, R. H., OLIVER, B., LIEB, J. D. & DISTECHE, C. M. 2011. Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nat Genet,* 43**,** 1179-85.

DESAI, S. D., LIU, L. F., VAZQUEZ-ABAD, D. & D'ARPA, P. 1997. Ubiquitin-dependent destruction of topoisomerase I is stimulated by the antitumor drug camptothecin. *J Biol Chem,* 272**,** 24159-64.

DEWING, P., CHIANG, C. W., SINCHAK, K., SIM, H., FERNAGUT, P. O., KELLY, S., CHESSELET, M. F., MICEVYCH, P. E., ALBRECHT, K. H., HARLEY, V. R. & VILAIN, E. 2006. Direct regulation of adult brain function by the male-specific factor SRY. *Curr Biol,* 16**,** 415-20.

DICARLO, J. E., NORVILLE, J. E., MALI, P., RIOS, X., AACH, J. & CHURCH, G. M. 2013. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res,* 41**,** 4336-43.

DIETRICH, J. E. & HIIRAGI, T. 2007. Stochastic patterning in the mouse pre-implantation embryo. *Development,* 134**,** 4219-31.

DISTECHE, C. M. & BERLETCH, J. B. 2015. X-chromosome inactivation and escape. *J Genet,* 94**,** 591-9.

DISTECHE, C. M., FILIPPOVA, G. N. & TSUCHIYA, K. D. 2002. Escape from X inactivation. *Cytogenetic and Genome Research,* 99**,** 36-43.

DISTECHE, C. M., ZACKSENHAUS, E., ADLER, D. A., BRESSLER, S. L., KEITZ, B. T. & CHAPMAN, V. M. 1992. Mapping and expression of the ubiquitin-activating enzyme E1 (Ube1) gene in the mouse. *Mamm Genome,* 3**,** 156-61.

DOENCH, J. G., FUSI, N., SULLENDER, M., HEGDE, M., VAIMBERG, E. W., DONOVAN, K. F., SMITH, I., TOTHOVA, Z., WILEN, C., ORCHARD, R., VIRGIN, H. W., LISTGARTEN, J. & ROOT, D. E. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology,* 34**,** 184.

DOLGANOV, G. M., MASER, R. S., NOVIKOV, A., TOSTO, L., CHONG, S., BRESSAN, D. A. & PETRINI, J. H. 1996. Human Rad50 is physically associated with human Mre11: identification of a conserved multiprotein complex implicated in recombinational DNA repair. *Mol Cell Biol,* 16**,** 4832-41.

DOUDNA, J. A. & CHARPENTIER, E. 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science,* 346**,** 1258096.

DUAN, J. E., SHI, W., JUE, N. K., JIANG, Z., KUO, L., O'NEILL, R., WOLF, E., DONG, H., ZHENG, X., CHEN, J. & TIAN, X. C. 2019. Dosage Compensation of the X Chromosomes in Bovine Germline, Early Embryos, and Somatic Tissues. *Genome Biol Evol,* 11**,** 242-252.

DUCIBELLA, T. & ANDERSON, E. 1975. Cell shape and membrane changes in the eight-cell mouse embryo: prerequisites for morphogenesis of the blastocyst. *Dev Biol,* 47**,** 45-58.

DUTTA, A., LE MAGNEN, C., MITROFANOVA, A., OUYANG, X., CALIFANO, A. & ABATE-SHEN, C. 2016. Identification of an NKX3.1-G9a-UTY transcriptional regulatory network that controls prostate differentiation. *Science,* 352**,** 1576-80.

ELLIS, N. & GOODFELLOW, P. N. 1989. The mammalian pseudoautosomal region. *Trends Genet,* 5**,** 406-10.

ENGREITZ, J. M., PANDYA-JONES, A., MCDONEL, P., SHISHKIN, A., SIROKMAN, K., SURKA, C., KADRI, S., XING, J., GOREN, A., LANDER, E. S., PLATH, K. & GUTTMAN, M. 2013. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science,* 341**,** 1237973.

ESCRIBANO-DIAZ, C., ORTHWEIN, A., FRADET-TURCOTTE, A., XING, M., YOUNG, J. T., TKAC, J., COOK, M. A., ROSEBROCK, A. P., MUNRO, M., CANNY, M. D., XU, D. & DUROCHER, D. 2013. A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol Cell,* 49**,** 872-83.

ESVELT, K. M., SMIDLER, A. L., CATTERUCCIA, F. & CHURCH, G. M. 2014. Concerning RNA-guided gene drives for the alteration of wild populations. *Elife,* 3.

FELSENSTEIN, J. 1974. The evolutionary advantage of recombination. *Genetics,* 78**,** 737-56.

FINEMAN, R. M., SCHOENWOLF, G. C., HUFF, M. & DAVIS, P. L. 1986. Causes of windowing-induced dysmorphogenesis (neural tube defects and early amnion deficit spectrum) in chicken embryos. *Am J Med Genet,* 25**,** 489-505.

FLACH, G., JOHNSON, M. H., BRAUDE, P. R., TAYLOR, R. A. & BOLTON, V. N. 1982. The transition from maternal to embryonic control in the 2-cell mouse embryo. *EMBO J,* 1**,** 681-6.

FORD, C. E., JONES, K. W., POLANI, P. E., DE ALMEIDA, J. C. & BRIGGS, J. H. 1959. A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet,* 1**,** 711-3.

FOSTER, J. W., BRENNAN, F. E., HAMPIKIAN, G. K., GOODFELLOW, P. N., SINCLAIR, A. H., LOVELL-BADGE, R., SELWOOD, L., RENFREE, M. B., COOPER, D. W. & GRAVES, J. A. 1992. Evolution of sex determination and the Y chromosome: SRY-related sequences in marsupials. *Nature,* 359**,** 531-3.

FRIDOLFSSON, A. K., CHENG, H., COPELAND, N. G., JENKINS, N. A., LIU, H. C., RAUDSEPP, T., WOODAGE, T., CHOWDHARY, B., HALVERSON, J. & ELLEGREN, H. 1998. Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proc Natl Acad Sci U S A,* 95**,** 8147-52.

FRIEDRICH, G. & SORIANO, P. 1991. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev,* 5**,** 1513-23.

FRIJTERS, A. C., MULLAART, E., ROELOFS, R. M., VAN HOORNE, R. P., MORENO, J. F., MORENO, O. & MERTON, J. S. 2009. What affects fertility of sexed bull semen more, low sperm dosage or the sorting process? *Theriogenology,* 71**,** 64-7.

GALIZI, R., DOYLE, L. A., MENICHELLI, M., BERNARDINI, F., DEREDEC, A., BURT, A., STODDARD, B. L., WINDBICHLER, N. & CRISANTI, A. 2014. A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nat Commun,* 5**,** 3977.

GALIZI, R., HAMMOND, A., KYROU, K., TAXIARCHI, C., BERNARDINI, F., O'LOUGHLIN, S. M., PAPATHANOS, P. A., NOLAN, T., WINDBICHLER, N. & CRISANTI, A. 2016. A CRISPR-Cas9 sex-ratio distortion system for genetic control. *Sci Rep,* 6**,** 31139.

GALLI, R., PREUSSE, G., SCHNABEL, C., BARTELS, T., CRAMER, K., KRAUTWALD-JUNGHANNS, M. E., KOCH, E. & STEINER, G. 2018. Sexing of chicken eggs by fluorescence and Raman spectroscopy through the shell membrane. *PLoS One,* 13**,** e0192554.

GALLI, R., PREUSSE, G., UCKERMANN, O., BARTELS, T., KRAUTWALD-JUNGHANNS, M. E., KOCH, E. & STEINER, G. 2016. In Ovo Sexing of Domestic Chicken Eggs by Raman Spectroscopy. *Anal Chem,* 88**,** 8657-63.

GALLI, R., PREUSSE, G., UCKERMANN, O., BARTELS, T., KRAUTWALD-JUNGHANNS, M. E., KOCH, E. & STEINER, G. 2017. In ovo sexing of chicken eggs by fluorescence spectroscopy. *Anal Bioanal Chem,* 409**,** 1185-1194.

GANTZ, V. M. & BIER, E. 2015. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science,* 348**,** 442-4.

GANTZ, V. M., JASINSKIENE, N., TATARENKOVA, O., FAZEKAS, A., MACIAS, V. M., BIER, E. & JAMES, A. A. 2015. Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proc Natl Acad Sci U S A,* 112**,** E6736-43.

GARDNER, R. L. 1985. Clonal analysis of early mammalian development. *Philos Trans R Soc Lond B Biol Sci,* 312**,** 163-78.

GARIERI, M., STAMOULIS, G., BLANC, X., FALCONNET, E., RIBAUX, P., BOREL, C., SANTONI, F. & ANTONARAKIS, S. E. 2018. Extensive cellular heterogeneity of X inactivation revealed by single-cell allele-specific expression in human fibroblasts. *Proceedings of the National Academy of Sciences,* 115**,** 13015.

GARNER, D. L. 2009. Hoechst 33342: the dye that enabled differentiation of living X- and Y-chromosome bearing mammalian sperm. *Theriogenology,* 71**,** 11-21.

GASIUNAS, G., BARRANGOU, R., HORVATH, P. & SIKSNYS, V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A,* 109**,** E2579-86.

GASPERINI, M., FINDLAY, G. M., MCKENNA, A., MILBANK, J. H., LEE, C., ZHANG, M. D., CUSANOVICH, D. A. & SHENDURE, J. 2017. CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. *Am J Hum Genet,* 101**,** 192-205.

GOHLER, D., FISCHER, B. & MEISSNER, S. 2017. In-ovo sexing of 14-day-old chicken embryos by pattern analysis in hyperspectral images (VIS/NIR spectra): A non-destructive method for layer lines with gender-specific down feather color. *Poult Sci,* 96**,** 1-4.

GONEN, N., FUTTNER, C. R., WOOD, S., GARCIA-MORENO, S. A., SALAMONE, I. M., SAMSON, S. C., SEKIDO, R., POULAT, F., MAATOUK, D. M. &

LOVELL-BADGE, R. 2018. Sex reversal following deletion of a single distal enhancer of Sox9. *Science,* 360**,** 1469.
GONEN, N., QUINN, A., O'NEILL, H. C., KOOPMAN, P. & LOVELL-BADGE, R. 2017. Normal Levels of Sox9 Expression in the Developing Mouse Testis Depend on the TES/TESCO Enhancer, but This Does Not Act Alone. *PLOS Genetics,* 13**,** e1006520.
GONG, S., YU, H. H., JOHNSON, K. A. & TAYLOR, D. W. 2018. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep,* 22**,** 359-371.
GOTO, Y. & TAKAGI, N. 2000. Maternally inherited X chromosome is not inactivated in mouse blastocysts due to parental imprinting. *Chromosome Res,* 8**,** 101-9.
GOTTLIEB, T. M. & JACKSON, S. P. 1993. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell,* 72**,** 131-42.
GRAVES, J. A. 1995. The evolution of mammalian sex chromosomes and the origin of sex determining genes. *Philos Trans R Soc Lond B Biol Sci,* 350**,** 305-11; discussion 311-2.
GRAVES, J. A. 2004. The degenerate Y chromosome--can conversion save it? *Reprod Fertil Dev,* 16**,** 527-34.
GRAVES, J. A. 2006. Sex chromosome specialization and degeneration in mammals. *Cell,* 124**,** 901-14.
GRAVES, J. A., KOINA, E. & SANKOVIC, N. 2006. How the gene content of human sex chromosomes evolved. *Curr Opin Genet Dev,* 16**,** 219-24.
GRAWUNDER, U., WILM, M., WU, X., KULESZA, P., WILSON, T. E., MANN, M. & LIEBER, M. R. 1997. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature,* 388**,** 492-5.
GREENFIELD, A., CARREL, L., PENNISI, D., PHILIPPE, C., QUADERI, N., SIGGERS, P., STEINER, K., TAM, P. P. L., MONACO, A. P., WILLARD, H. F. & KOOPMAN, P. 1998. The UTX Gene Escapes X Inactivation in Mice and Humans. *Human Molecular Genetics,* 7**,** 737-742.
GREENFIELD, A., SCOTT, D., PENNISI, D., EHRMANN, I., ELLIS, P., COOPER, L., SIMPSON, E. & KOOPMAN, P. 1996. An H–YDb epitope is encoded by a novel mouse Y chromosome gene. *Nature Genetics,* 14**,** 474-478.
GRUMBACH, M. M., MORISHIMA, A. & TAYLOR, J. H. 1963. Human Sex Chromosome Abnormalities in Relation to DNA Replication and Heterochromatinization. *Proc Natl Acad Sci U S A,* 49**,** 581-9.
GRUNDY, G. J., RULTEN, S. L., ZENG, Z., ARRIBAS-BOSACOMA, R., ILES, N., MANLEY, K., OLIVER, A. & CALDECOTT, K. W. 2013. APLF promotes the assembly and activity of non-homologous end joining protein complexes. *EMBO J,* 32**,** 112-25.
GRUNWALD, H. A., GANTZ, V. M., POPLAWSKI, G., XU, X. S., BIER, E. & COOPER, K. L. 2019. Super-Mendelian inheritance mediated by CRISPR-Cas9 in the female mouse germline. *Nature,* 566**,** 105-109.
GU, J., LU, H., TIPPIN, B., SHIMAZAKI, N., GOODMAN, M. F. & LIEBER, M. R. 2007a. XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO J,* 26**,** 1010-23.
GU, J., LU, H., TSAI, A. G., SCHWARZ, K. & LIEBER, M. R. 2007b. Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. *Nucleic Acids Res,* 35**,** 5755-62.

GUBBAY, J., COLLIGNON, J., KOOPMAN, P., CAPEL, B., ECONOMOU, A., MUNSTERBERG, A., VIVIAN, N., GOODFELLOW, P. & LOVELL-BADGE, R. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature,* 346**,** 245-50.

GUPTA, V., PARISI, M., STURGILL, D., NUTTALL, R., DOCTOLERO, M., DUDKO, O. K., MALLEY, J. D., EASTMAN, P. S. & OLIVER, B. 2006. Global analysis of X-chromosome dosage compensation. *J Biol,* 5**,** 3.

HAEUSSLER, M., SCHONIG, K., ECKERT, H., ESCHSTRUTH, A., MIANNE, J., RENAUD, J. B., SCHNEIDER-MAUNOURY, S., SHKUMATAVA, A., TEBOUL, L., KENT, J., JOLY, J. S. & CONCORDET, J. P. 2016. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol,* 17**,** 148.

HAFT, D. H., SELENGUT, J., MONGODIN, E. F. & NELSON, K. E. 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol,* 1**,** e60.

HAMILTON, W. D. 1967. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science,* 156**,** 477-88.

HAMMOND, A., GALIZI, R., KYROU, K., SIMONI, A., SINISCALCHI, C., KATSANOS, D., GRIBBLE, M., BAKER, D., MAROIS, E., RUSSELL, S., BURT, A., WINDBICHLER, N., CRISANTI, A. & NOLAN, T. 2016. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat Biotechnol,* 34**,** 78-83.

HAMMOND, A. M. & GALIZI, R. 2017. Gene drives to fight malaria: current state and future directions. *Pathog Glob Health,* 111**,** 412-423.

HAMMOND, A. M., KYROU, K., BRUTTINI, M., NORTH, A., GALIZI, R., KARLSSON, X., KRANJC, N., CARPI, F. M., D'AURIZIO, R., CRISANTI, A. & NOLAN, T. 2017. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet,* 13**,** e1007039.

HARLEY, V. R. & GOODFELLOW, P. N. 1994. The biochemical role of SRY in sex determination. *Mol Reprod Dev,* 39**,** 184-93.

HARLEY, V. R., LOVELL-BADGE, R. & GOODFELLOW, P. N. 1994. Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res,* 22**,** 1500-1.

HATTORI, R. S., MURAI, Y., OURA, M., MASUDA, S., MAJHI, S. K., SAKAMOTO, T., FERNANDINO, J. I., SOMOZA, G. M., YOKOTA, M. & STRUSSMANN, C. A. 2012. A Y-linked anti-Mullerian hormone duplication takes over a critical role in sex determination. *Proc Natl Acad Sci U S A,* 109**,** 2955-9.

HAYASHI, K. & SAITOU, M. 2013. Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells. *Nature Protocols,* 8**,** 1513.

HEARD, E., CLERC, P. & AVNER, P. 1997. X-chromosome inactivation in mammals. *Annu Rev Genet,* 31**,** 571-610.

HEARD, E., ROUGEULLE, C., ARNAUD, D., AVNER, P., ALLIS, C. D. & SPECTOR, D. L. 2001. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell,* 107**,** 727-38.

HIPPENMEYER, S., YOUN, Y. H., MOON, H. M., MIYAMICHI, K., ZONG, H., WYNSHAW-BORIS, A. & LUO, L. 2010. Genetic mosaic dissection of Lis1 and Ndel1 in neuronal migration. *Neuron,* 68**,** 695-709.

HOLLIDAY, R. 2007. A mechanism for gene conversion in fungi. *Genet Res,* 89**,** 285-307.

HOOPER, M., HARDY, K., HANDYSIDE, A., HUNTER, S. & MONK, M. 1987. HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature,* 326**,** 292-295.

HSU, H. L., YANNONE, S. M. & CHEN, D. J. 2002. Defining interactions between DNA-PK and ligase IV/XRCC4. *DNA Repair (Amst),* 1**,** 225-35.

HUGHES, J. F., SKALETSKY, H., BROWN, L. G., PYNTIKOVA, T., GRAVES, T., FULTON, R. S., DUGAN, S., DING, Y., BUHAY, C. J., KREMITZKI, C., WANG, Q., SHEN, H., HOLDER, M., VILLASANA, D., NAZARETH, L. V., CREE, A., COURTNEY, L., VEIZER, J., KOTKIEWICZ, H., CHO, T. J., KOUTSEVA, N., ROZEN, S., MUZNY, D. M., WARREN, W. C., GIBBS, R. A., WILSON, R. K. & PAGE, D. C. 2012. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature,* 483**,** 82-6.

HUGHES, J. F., SKALETSKY, H., PYNTIKOVA, T., GRAVES, T. A., VAN DAALEN, S. K., MINX, P. J., FULTON, R. S., MCGRATH, S. D., LOCKE, D. P., FRIEDMAN, C., TRASK, B. J., MARDIS, E. R., WARREN, W. C., REPPING, S., ROZEN, S., WILSON, R. K. & PAGE, D. C. 2010. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature,* 463**,** 536-9.

HUGHES, J. F., SKALETSKY, H., PYNTIKOVA, T., MINX, P. J., GRAVES, T., ROZEN, S., WILSON, R. K. & PAGE, D. C. 2005. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature,* 437**,** 100-3.

HURST, L. D. 2001. Evolutionary genomics. Sex and the X. *Nature,* 411**,** 149-50.

HUYNH, K. D. & LEE, J. T. 2003. Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. *Nature,* 426**,** 857-62.

IMAIMATSU, K., FUJII, W., HIRAMATSU, R., MIURA, K., KUROHMARU, M. & KANAI, Y. 2018. CRISPR/Cas9-mediated knock-in of the murine Y chromosomal Sry gene. *J Reprod Dev,* 64**,** 283-287.

INOUE, A., JIANG, L., LU, F., SUZUKI, T. & ZHANG, Y. 2017a. Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature,* 547**,** 419-424.

INOUE, A., JIANG, L., LU, F. & ZHANG, Y. 2017b. Genomic imprinting of Xist by maternal H3K27me3. *Genes Dev,* 31**,** 1927-1932.

ISHINO, Y., SHINAGAWA, H., MAKINO, K., AMEMURA, M. & NAKATA, A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol,* 169**,** 5429-33.

JACOBS, P. A. & STRONG, J. A. 1959. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature,* 183**,** 302-3.

JANSEN, R., EMBDEN, J. D., GAASTRA, W. & SCHOULS, L. M. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol,* 43**,** 1565-75.

JASIN, M., DE VILLIERS, J., WEBER, F. & SCHAFFNER, W. 1985. High frequency of homologous recombination in mammalian cells between endogenous and introduced SV40 genomes. *Cell,* 43**,** 695-703.

JASIN, M. & ROTHSTEIN, R. 2013. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol,* 5**,** a012740.

JAYATHILAKA, K., SHERIDAN, S. D., BOLD, T. D., BOCHENSKA, K., LOGAN, H. L., WEICHSELBAUM, R. R., BISHOP, D. K. & CONNELL, P. P. 2008. A chemical compound that stimulates the human homologous recombination protein RAD51. *Proc Natl Acad Sci U S A,* 105**,** 15848-53.

JESKE, Y. W. A., BOWLES, J., GREENFIELD, A. & KOOPMAN, P. 1995. Expression of a linear Sry transcript in the mouse genital ridge. *Nature Genetics,* 10**,** 480-482.

JIANG, Y., XIE, M., CHEN, W., TALBOT, R., MADDOX, J. F., FARAUT, T., WU, C., MUZNY, D. M., LI, Y., ZHANG, W., STANTON, J. A., BRAUNING, R., BARRIS, W. C., HOURLIER, T., AKEN, B. L., SEARLE, S. M. J., ADELSON, D. L., BIAN, C., CAM, G. R., CHEN, Y., CHENG, S., DESILVA, U., DIXEN, K., DONG, Y., FAN, G., FRANKLIN, I. R., FU, S., GUAN, R., HIGHLAND, M. A., HOLDER, M. E., HUANG, G., INGHAM, A. B., JHANGIANI, S. N., KALRA, D., KOVAR, C. L., LEE, S. L., LIU, W., LIU, X., LU, C., LV, T., MATHEW, T., MCWILLIAM, S., MENZIES, M., PAN, S., ROBELIN, D., SERVIN, B., TOWNLEY, D., WANG, W., WEI, B., WHITE, S. N., YANG, X., YE, C., YUE, Y., ZENG, P., ZHOU, Q., HANSEN, J. B., KRISTENSEN, K., GIBBS, R. A., FLICEK, P., WARKUP, C. C., JONES, H. E., ODDY, V. H., NICHOLAS, F. W., MCEWAN, J. C., KIJAS, J., WANG, J., WORLEY, K. C., ARCHIBALD, A. L., COCKETT, N., XU, X., WANG, W. & DALRYMPLE, B. P. 2014. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science,* 344**,** 1168-1173.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337**,** 816-21.

JINEK, M., EAST, A., CHENG, A., LIN, S., MA, E. & DOUDNA, J. 2013. RNA-programmed genome editing in human cells. *Elife,* 2**,** e00471.

JINEK, M., JIANG, F., TAYLOR, D. W., STERNBERG, S. H., KAYA, E., MA, E., ANDERS, C., HAUER, M., ZHOU, K., LIN, S., KAPLAN, M., IAVARONE, A. T., CHARPENTIER, E., NOGALES, E. & DOUDNA, J. A. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science,* 343**,** 1247997.

JINNAH, H. A., GAGE, F. H. & FRIEDMANN, T. 1990. Animal models of Lesch-Nyhan syndrome. *Brain Research Bulletin,* 25**,** 467-475.

JOHNSON, L. A. 1995. Sex preselection by flow cytometric separation of X and Y chromosome-bearing sperm based on DNA difference: a review. *Reprod Fertil Dev,* 7**,** 893-903.

JOHNSON, L. A. & CLARKE, R. N. 1988. Flow sorting of X and Y chromosome-bearing mammalian sperm: activation and pronuclear development of sorted bull, boar, and ram sperm microinjected into hamster oocytes. *Gamete Res,* 21**,** 335-43.

JOHNSON, L. A., FLOOK, J. P. & HAWK, H. W. 1989. Sex preselection in rabbits: live births from X and Y sperm separated by DNA and cell sorting. *Biol Reprod,* 41**,** 199-203.

JOHNSON, L. A., FLOOK, J. P. & LOOK, M. V. 1987. Flow cytometry of X and Y chromosome-bearing sperm for DNA using an improved preparation method and staining with Hoechst 33342. *Gamete Res,* 17**,** 203-12.

JOHNSON, L. A. & PINKEL, D. 1986. Modification of a laser-based flow cytometer for high-resolution DNA analysis of mammalian spermatozoa. *Cytometry,* 7**,** 268-73.

JOHNSON, L. A., WELCH, G. R. & RENS, W. 1999. The Beltsville sperm sexing technology: high-speed sperm sorting gives improved sperm output for in vitro fertilization and AI. *J Anim Sci,* 77 Suppl 2**,** 213-20.

JONKERS, I., BARAKAT, T. S., ACHAME, E. M., MONKHORST, K., KENTER, A., RENTMEESTER, E., GROSVELD, F., GROOTEGOED, J. A. & GRIBNAU, J. 2009. RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. *Cell,* 139**,** 999-1011.

JULIEN, P., BRAWAND, D., SOUMILLON, M., NECSULEA, A., LIECHTI, A., SCHUTZ, F., DAISH, T., GRUTZNER, F. & KAESSMANN, H. 2012. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol,* 10**,** e1001328.

KA, S., AHN, H., SEO, M., KIM, H., KIM, J. N. & LEE, H. J. 2016. Status of dosage compensation of X chromosome in bovine genome. *Genetica,* 144**,** 435-44.

KAMACHI, Y. & KONDOH, H. 2013. Sox proteins: regulators of cell fate specification and differentiation. *Development,* 140**,** 4129-44.

KASHIMADA, K. & KOOPMAN, P. 2010. Sry: the master switch in mammalian sex determination. *Development,* 137**,** 3921-30.

KASS, E. M. & JASIN, M. 2010. Collaboration and competition between DNA double-strand break repair pathways. *FEBS Lett,* 584**,** 3703-8.

KAWARASAKI, T., WELCH, G. R., LONG, C. R., YOSHIDA, M. & JOHNSON, L. A. 1998. Verification of flow cytometorically-sorted X- and Y-bearing porcine spermatozoa and reanalysis of spermatozoa for DNA content using the fluorescence in situ hybridization (FISH) technique. *Theriogenology,* 50**,** 625-35.

KAY, G. F., PENNY, G. D., PATEL, D., ASHWORTH, A., BROCKDORFF, N. & RASTAN, S. 1993. Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. *Cell,* 72**,** 171-82.

KHARCHENKO, P. V., XI, R. & PARK, P. J. 2011. Evidence for dosage compensation between the X chromosome and autosomes in mammals. *Nat Genet,* 43**,** 1167-9; author reply 1171-2.

KHIL, P. P., SMIRNOVA, N. A., ROMANIENKO, P. J. & CAMERINI-OTERO, R. D. 2004. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nat Genet,* 36**,** 642-6.

KIM, D., LANGMEAD, B. & SALZBERG, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods,* 12**,** 357-60.

KITA, M. & IMAI, H. 1993. Hypoxanthine phosphoribosyltransferase activity in bovine embryos during the early embryonic development. *Theriogenology,* 40**,** 357-364.

KOBAYASHI, M., AIDA, M., NAGAOKA, H., BEGUM, N. A., KITAWAKI, Y., NAKATA, M., STANLIE, A., DOI, T., KATO, L., OKAZAKI, I. M., SHINKURA, R., MURAMATSU, M., KINOSHITA, K. & HONJO, T. 2009. AID-induced decrease in topoisomerase 1 induces DNA structural alteration and DNA cleavage for class switch recombination. *Proc Natl Acad Sci U S A,* 106**,** 22375-80.

KOBAYASHI, M., SABOURI, Z., SABOURI, S., KITAWAKI, Y., POMMIER, Y., ABE, T., KIYONARI, H. & HONJO, T. 2011. Decrease in topoisomerase I is responsible for activation-induced cytidine deaminase (AID)-dependent somatic hypermutation. *Proc Natl Acad Sci U S A,* 108**,** 19305-10.

KOBAYASHI, T., KATO-ITOH, M., YAMAGUCHI, T., TAMURA, C., SANBO, M., HIRABAYASHI, M. & NAKAUCHI, H. 2012. Identification of rat Rosa26 locus enables generation of knock-in rat lines ubiquitously expressing tdTomato. *Stem cells and development,* 21**,** 2981-2986.

KOHNO, S., PARROTT, B. B., YATSU, R., MIYAGAWA, S., MOORE, B. C., IGUCHI, T. & GUILLETTE, L., JR. 2014. Gonadal differentiation in reptiles exhibiting environmental sex determination. *Sex Dev,* 8**,** 208-26.

KOLLER, B. H., HAGEMANN, L. J., DOETSCHMAN, T., HAGAMAN, J. R., HUANG, S., WILLIAMS, P. J., FIRST, N. L., MAEDA, N. & SMITHIES, O. 1989. Germ-line transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America,* 86**,** 8927-8931.

KONG, Q., HAI, T., MA, J., HUANG, T., JIANG, D., XIE, B., WU, M., WANG, J., SONG, Y., WANG, Y., HE, Y., SUN, J., HU, K., GUO, R., WANG, L., ZHOU, Q., MU, Y. & LIU, Z. 2014. Rosa26 locus supports tissue-specific promoter driving transgene expression specifically in pig. *PloS one,* 9**,** e107945-e107945.

KOOPMAN, P., GUBBAY, J., VIVIAN, N., GOODFELLOW, P. & LOVELL-BADGE, R. 1991. Male development of chromosomally female mice transgenic for Sry. *Nature,* 351**,** 117-21.

KOOPMAN, P., MUNSTERBERG, A., CAPEL, B., VIVIAN, N. & LOVELL-BADGE, R. 1990. Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature,* 348**,** 450-2.

KORESSAAR, T., LEPAMETS, M., KAPLINSKI, L., RAIME, K., ANDRESON, R. & REMM, M. 2018. Primer3_masker: integrating masking of template sequence with primer design software. *Bioinformatics,* 34**,** 1937-1938.

KORESSAAR, T. & REMM, M. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics,* 23**,** 1289-91.

KOSICKI, M., TOMBERG, K. & BRADLEY, A. 2018. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol,* 36**,** 765-771.

KRAFT, K., GEUER, S., WILL, A. J., CHAN, W. L., PALIOU, C., BORSCHIWER, M., HARABULA, I., WITTLER, L., FRANKE, M., IBRAHIM, D. M., KRAGESTEEN, B. K., SPIELMANN, M., MUNDLOS, S., LUPIANEZ, D. G. & ANDREY, G. 2015. Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell Rep,* 10**,** 833-839.

KRAUTWALD-JUNGHANNS, M. E., CRAMER, K., FISCHER, B., FORSTER, A., GALLI, R., KREMER, F., MAPESA, E. U., MEISSNER, S., PREISINGER, R., PREUSSE, G., SCHNABEL, C., STEINER, G. & BARTELS, T. 2018. Current approaches to avoid the culling of day-old male chicks in the layer industry, with special reference to spectroscopic methods. *Poult Sci,* 97**,** 749-757.

KUEHN, M. R., BRADLEY, A., ROBERTSON, E. J. & EVANS, M. J. 1987. A potential animal model for Lesch–Nyhan syndrome through introduction of HPRT mutations into mice. *Nature,* 326**,** 295-298.

KWART, D., PAQUET, D., TEO, S. & TESSIER-LAVIGNE, M. 2017. Precise and efficient scarless genome editing in stem cells using CORRECT. *Nat Protoc,* 12**,** 329-354.

LAHN, B. T. & PAGE, D. C. 1997. Functional coherence of the human Y chromosome. *Science,* 278**,** 675-80.

LALLEMAND, Y., LURIA, V., HAFFNER-KRAUSZ, R. & LONAI, P. 1998. Maternally expressed PGK-Cre transgene as a tool for early and uniform activation of the Cre site-specific recombinase. *Transgenic Res,* 7**,** 105-12.

LEE, J. T. 2002. Homozygous Tsix mutant mice reveal a sex-ratio distortion and revert to random X-inactivation. *Nat Genet,* 32**,** 195-200.

LEE, J. T. 2005. Regulation of X-chromosome counting by Tsix and Xite sequences. *Science,* 309**,** 768-71.

LEE, J. T., DAVIDOW, L. S. & WARSHAWSKY, D. 1999. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet,* 21**,** 400-4.

LEE, J. T. & JAENISCH, R. 1997. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. *Nature,* 386**,** 275-9.

LEE, J. T. & LU, N. 1999. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell,* 99**,** 47-57.

LEE, J. T., STRAUSS, W. M., DAUSMAN, J. A. & JAENISCH, R. 1996. A 450 kb transgene displays properties of the mammalian X-inactivation center. *Cell,* 86**,** 83-94.

LEJEUNE, J., LEVAN, A., BÖÖK, J. A., CHU, E. H. Y., FORD, C. E., FRACCARO, M., HARNDEN, D. G., HSU, T. C., HUNGERFORD, D. A., JACOBS, P. A., MAKINO, S., PUCK, T., ROBINSON, A., TJIO, J. H., CATCHESIDE, D. G., MULLER, H. J. & STERN, C. 1960. A PROPOSED STANDARD SYSTEM OF NOMENCLATURE OF HUMAN MITOTIC CHROMOSOMES. *The Lancet,* 275**,** 1063-1065.

LEPPIG, K. A., BROWN, C. J., BRESSLER, S. L., GUSTASHAW, K., PAGON, R. A., WILLARD, H. F. & DISTECHE, C. M. 1993. Mapping of the distal boundary of the X-inactivation center in a rearranged X chromosome from a female expressing XIST. *Hum Mol Genet,* 2**,** 883-7.

LERCHER, M. J., URRUTIA, A. O. & HURST, L. D. 2003. Evidence that the human X chromosome is enriched for male-specific but not female-specific genes. *Mol Biol Evol,* 20**,** 1113-6.

LESCH, M. & NYHAN, W. L. 1964. A familial disorder of uric acid metabolism and central nervous system function. *The American Journal of Medicine,* 36**,** 561-570.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, S., CHANG, H. H., NIEWOLIK, D., HEDRICK, M. P., PINKERTON, A. B., HASSIG, C. A., SCHWARZ, K. & LIEBER, M. R. 2014a. Evidence that the DNA endonuclease ARTEMIS also has intrinsic 5'-exonuclease activity. *J Biol Chem,* 289**,** 7825-34.

LI, S., FLISIKOWSKA, T., KUROME, M., ZAKHARTCHENKO, V., KESSLER, B., SAUR, D., KIND, A., WOLF, E., FLISIKOWSKI, K. & SCHNIEKE, A. 2014b. Dual fluorescent reporter pig for Cre recombination: transgene placement at the ROSA26 locus. *PloS one,* 9**,** e102455-e102455.

LI, X., YANG, Y., BU, L., GUO, X., TANG, C., SONG, J., FAN, N., ZHAO, B., OUYANG, Z., LIU, Z., ZHAO, Y., YI, X., QUAN, L., LIU, S., YANG, Z., OUYANG, H., CHEN, Y. E., WANG, Z. & LAI, L. 2014c. Rosa26-targeted swine models for stable gene over-expression and Cre-mediated lineage tracing. *Cell research,* 24**,** 501-504.

LIAO, Y., SMYTH, G. K. & SHI, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* 30**,** 923-30.

LIEBER, M. R. 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem,* 79**,** 181-211.

LIN, F., XING, K., ZHANG, J. & HE, X. 2012. Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation. *Proc Natl Acad Sci U S A,* 109**,** 11752-7.

LIN, H., HALSALL, J. A., ANTCZAK, P., O'NEILL, L. P., FALCIANI, F. & TURNER, B. M. 2011. Relative overexpression of X-linked genes in mouse embryonic stem cells is consistent with Ohno's hypothesis. *Nature Genetics,* 43**,** 1169.

LIN, S., STAAHL, B. T., ALLA, R. K. & DOUDNA, J. A. 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife,* 3**,** e04766.

LINDSAY, H., BURGER, A., BIYONG, B., FELKER, A., HESS, C., ZAUGG, J., CHIAVACCI, E., ANDERS, C., JINEK, M., MOSIMANN, C. & ROBINSON, M. D. 2016. CrispRVariants charts the mutation spectrum of genome engineering experiments. *Nat Biotechnol,* 34**,** 701-2.

LISTER, R., MUKAMEL, E. A., NERY, J. R., URICH, M., PUDDIFOOT, C. A., JOHNSON, N. D., LUCERO, J., HUANG, Y., DWORK, A. J., SCHULTZ, M. D., YU, M., TONTI-FILIPPINI, J., HEYN, H., HU, S., WU, J. C., RAO, A., ESTELLER, M., HE, C., HAGHIGHI, F. G., SEJNOWSKI, T. J., BEHRENS, M. M. & ECKER, J. R. 2013. Global epigenomic reconfiguration during mammalian brain development. *Science,* 341**,** 1237905.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

LOVELL-BADGE, R., CANNING, C. & SEKIDO, R. 2002. Sex-determining genes in mice: building pathways. *Novartis Found Symp,* 244**,** 4-18; discussion 18-22, 35-42, 253-7.

LYON, M. F. 1961. Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature,* 190**,** 372-3.

LYON, M. F. 1962. Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet,* 14**,** 135-48.

LYON, M. F. 1988. The William Allan memorial award address: X-chromosome inactivation and the location and expression of X-linked genes. *Am J Hum Genet,* 42**,** 8-16.

LYON, M. F. 1998. X-chromosome inactivation: a repeat hypothesis. *Cytogenet Cell Genet,* 80**,** 133-7.

LYON, M. F. 2003. Transmission ratio distortion in mice. *Annu Rev Genet,* 37**,** 393-408.

MA, Y., CHEN, W., ZHANG, X., YU, L., DONG, W., PAN, S., GAO, S., HUANG, X. & ZHANG, L. 2016. Increasing the efficiency of CRISPR/Cas9-mediated precise genome editing in rats by inhibiting NHEJ and using Cas9 protein. *RNA Biol,* 13**,** 605-12.

MABB, A. M., SIMON, J. M., KING, I. F., LEE, H. M., AN, L. K., PHILPOT, B. D. & ZYLKA, M. J. 2016. Topoisomerase 1 Regulates Gene Expression in Neurons through Cleavage Complex-Dependent and -Independent Mechanisms. *PLoS One,* 11**,** e0156439.

MADDALO, D., MANCHADO, E., CONCEPCION, C. P., BONETTI, C., VIDIGAL, J. A., HAN, Y. C., OGRODOWSKI, P., CRIPPA, A., REKHTMAN, N., DE STANCHINA, E., LOWE, S. W. & VENTURA, A. 2014. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature,* 516**,** 423-7.

MADISEN, L., ZWINGMAN, T. A., SUNKIN, S. M., OH, S. W., ZARIWALA, H. A., GU, H., NG, L. L., PALMITER, R. D., HAWRYLYCZ, M. J., JONES, A. R., LEIN, E. S. & ZENG, H. 2010. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci,* 13**,** 133-40.

MAK, W., NESTEROVA, T. B., DE NAPOLES, M., APPANAH, R., YAMANAKA, S., OTTE, A. P. & BROCKDORFF, N. 2004. Reactivation of the paternal X chromosome in early mouse embryos. *Science,* 303**,** 666-9.

MAKAROVA, K. S., GRISHIN, N. V., SHABALINA, S. A., WOLF, Y. I. & KOONIN, E. V. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct,* 1**,** 7.

MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013. RNA-guided human genome engineering via Cas9. *Science,* 339**,** 823-6.

MALIVERT, L., ROPARS, V., NUNEZ, M., DREVET, P., MIRON, S., FAURE, G., GUEROIS, R., MORNON, J. P., REVY, P., CHARBONNIER, J. B., CALLEBAUT, I. & DE VILLARTAY, J. P. 2010. Delineation of the Xrcc4-interacting region in the globular head domain of cernunnos/XLF. *J Biol Chem,* 285**,** 26475-83.

MARAHRENS, Y., PANNING, B., DAUSMAN, J., STRAUSS, W. & JAENISCH, R. 1997. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev,* 11**,** 156-66.

MARI, P. O., FLOREA, B. I., PERSENGIEV, S. P., VERKAIK, N. S., BRUGGENWIRTH, H. T., MODESTI, M., GIGLIA-MARI, G., BEZSTAROSTI, K., DEMMERS, J. A., LUIDER, T. M., HOUTSMULLER, A. B. & VAN GENT, D. C. 2006. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc Natl Acad Sci U S A,* 103**,** 18597-602.

MARKS, H., CHOW, J. C., DENISSOV, S., FRANCOIJS, K. J., BROCKDORFF, N., HEARD, E. & STUNNENBERG, H. G. 2009. High-resolution analysis of epigenetic changes associated with X inactivation. *Genome Res,* 19**,** 1361-73.

MARKS, H., KERSTENS, H. H., BARAKAT, T. S., SPLINTER, E., DIRKS, R. A., VAN MIERLO, G., JOSHI, O., WANG, S. Y., BABAK, T., ALBERS, C. A., KALKAN, T., SMITH, A., JOUNEAU, A., DE LAAT, W., GRIBNAU, J. & STUNNENBERG, H. G. 2015. Dynamics of gene silencing during X inactivation using allele-specific RNA-seq. *Genome Biol,* 16**,** 149.

MARQUES, R., WILLIAMS, A., EKSMOND, U., WULLAERT, A., KILLEEN, N., PASPARAKIS, M., KIOUSSIS, D. & KASSIOTIS, G. 2009. Generalized immune activation as a direct result of activated CD4+ T cell killing. *J Biol,* 8**,** 93.

MARRAFFINI, L. A. & SONTHEIMER, E. J. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science,* 322**,** 1843-5.

MARUYAMA, T., DOUGAN, S. K., TRUTTMANN, M. C., BILATE, A. M., INGRAM, J. R. & PLOEGH, H. L. 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol,* 33**,** 538-42.

MATSUBARA, Y., KATO, T., KASHIMADA, K., TANAKA, H., ZHI, Z., ICHINOSE, S., MIZUTANI, S., MORIO, T., CHIBA, T., ITO, Y., SAGA, Y., TAKADA, S. & ASAHARA, H. 2015. TALEN-Mediated Gene Disruption on Y Chromosome Reveals Critical Role of EIF2S3Y in Mouse Spermatogenesis. *Stem Cells Dev,* 24**,** 1164-70.

MATSUDA, Y., MOENS, P. B. & CHAPMAN, V. M. 1992. Deficiency of X and Y chromosomal pairing at meiotic prophase in spermatocytes of sterile interspecific hybrids between laboratory mice (Mus domesticus) and Mus spretus. *Chromosoma,* 101**,** 483-92.

MAZEYRAT, S., SAUT, N., GRIGORIEV, V., MAHADEVAIAH, S. K., OJARIKRE, O. A., RATTIGAN, A., BISHOP, C., EICHER, E. M., MITCHELL, M. J. & BURGOYNE, P. S. 2001. A Y-encoded subunit of the translation initiation factor Eif2 is essential for mouse spermatogenesis. *Nat Genet,* 29**,** 49-53.

MEEK, S., THOMSON, A. J., SUTHERLAND, L., SHARP, M. G. F., THOMSON, J., BISHOP, V., MEDDLE, S. L., GLOAGUEN, Y., WEIDT, S., SINGH-DOLT, K., BUEHR, M., BROWN, H. K., GILL, A. C. & BURDON, T. 2016. Reduced levels of dopamine and altered metabolism in brains of HPRT knock-out rats: a new rodent model of Lesch-Nyhan Disease. *Scientific Reports,* 6**,** 25592.

MIGEON, B. R., LEE, C. H., CHOWDHURY, A. K. & CARPENTER, H. 2002. Species differences in TSIX/Tsix reveal the roles of these genes in X-chromosome inactivation. *Am J Hum Genet,* 71**,** 286-93.

MIMITOU, E. P. & SYMINGTON, L. S. 2008. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature,* 455**,** 770-4.

MITSUI, K., TOKUZAWA, Y., ITOH, H., SEGAWA, K., MURAKAMI, M., TAKAHASHI, K., MARUYAMA, M., MAEDA, M. & YAMANAKA, S. 2003. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell,* 113**,** 631-42.

MOHANDAS, T. K., SPEED, R. M., PASSAGE, M. B., YEN, P. H., CHANDLEY, A. C. & SHAPIRO, L. J. 1992. Role of the pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion of distal Xp. *Am J Hum Genet,* 51**,** 526-33.

MOJICA, F. J., DIEZ-VILLASENOR, C., GARCIA-MARTINEZ, J. & SORIA, E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol,* 60**,** 174-82.

MOREY, C., ARNAUD, D., AVNER, P. & CLERC, P. 2001. Tsix-mediated repression of Xist accumulation is not sufficient for normal random X inactivation. *Hum Mol Genet,* 10**,** 1403-11.

MORHAM, S. G., KLUCKMAN, K. D., VOULOMANOS, N. & SMITHIES, O. 1996. Targeted disruption of the mouse topoisomerase I gene by camptothecin selection. *Mol Cell Biol,* 16**,** 6804-9.

MUELLER, J. L., MAHADEVAIAH, S. K., PARK, P. J., WARBURTON, P. E., PAGE, D. C. & TURNER, J. M. 2008. The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat Genet,* 40**,** 794-9.

MUELLER, J. L., SKALETSKY, H., BROWN, L. G., ZAGHLUL, S., ROCK, S., GRAVES, T., AUGER, K., WARREN, W. C., WILSON, R. K. & PAGE, D. C. 2013. Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet,* 45**,** 1083-7.

MULAS, C., KALKAN, T., VON MEYENN, F., LEITCH, H. G., NICHOLS, J. & SMITH, A. 2019a. Correction: Defined conditions for propagation and manipulation of mouse embryonic stem cells (doi:10.1242/dev.173146). *Development,* 146.

MULAS, C., KALKAN, T., VON MEYENN, F., LEITCH, H. G., NICHOLS, J. & SMITH, A. 2019b. Defined conditions for propagation and manipulation of mouse embryonic stem cells. *Development,* 146.

MULLER, H. J. 1914. A Factor for the Fourth Chromosome of Drosophila. *Science,* 39**,** 906.

MULLER, H. J. 1932. Some Genetic Aspects of Sex. *The American Naturalist,* 66**,** 118-138.

MULLER, H. J. 1964. The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 1**,** 2-9.

NADEAU, J. H. 2017. Do Gametes Woo? Evidence for Their Nonrandom Union at Fertilization. *Genetics,* 207**,** 369-387.

NAKOUZI, G. A. & NADEAU, J. H. 2014. Does dietary folic acid supplementation in mouse NTD models affect neural tube development or gamete preference at fertilization? *BMC genetics,* 15**,** 91-91.

NANDA, I., SHAN, Z., SCHARTL, M., BURT, D. W., KOEHLER, M., NOTHWANG, H., GRUTZNER, F., PATON, I. R., WINDSOR, D., DUNN, I., ENGEL, W., STAEHELI, P., MIZUNO, S., HAAF, T. & SCHMID, M. 1999. 300 million years of conserved synteny between chicken Z and human chromosome 9. *Nat Genet,* 21**,** 258-9.

NGUYEN, D. K. & DISTECHE, C. M. 2006. Dosage compensation of the active X chromosome in mammals. *Nat Genet,* 38**,** 47-53.

NICHOLS, J., ZEVNIK, B., ANASTASSIADIS, K., NIWA, H., KLEWE-NEBENIUS, D., CHAMBERS, I., SCHÖLER, H. & SMITH, A. 1998. Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell,* 95**,** 379-391.

NICK MCELHINNY, S. A. & RAMSDEN, D. A. 2003. Polymerase mu is a DNA-directed DNA/RNA polymerase. *Mol Cell Biol,* 23**,** 2309-15.

NISHIMASU, H., RAN, F. A., HSU, P. D., KONERMANN, S., SHEHATA, S. I., DOHMAE, N., ISHITANI, R., ZHANG, F. & NUREKI, O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell,* 156**,** 935-49.

NIWA, H., TOYOOKA, Y., SHIMOSATO, D., STRUMPF, D., TAKAHASHI, K., YAGI, R. & ROSSANT, J. 2005. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell,* 123**,** 917-29.

OHHATA, T., HOKI, Y., SASAKI, H. & SADO, T. 2006. Tsix-deficient X chromosome does not undergo inactivation in the embryonic lineage in males: implications for Tsix-independent silencing of Xist. *Cytogenet Genome Res,* 113**,** 345-9.

OHNO, S. 1967. *Sex chromosomes and sex-linked genes. (Monographs on endocrinology, Vol. 1.)*, Berlin, Heidelberg, New York: Springer Verlag.

OHNO, S., KAPLAN, W. D. & KINOSITA, R. 1959. Formation of the sex chromatin by a single X-chromosome in liver cells of Rattus norvegicus. *Exp Cell Res,* 18**,** 415-8.

OKAMOTO, I., ARNAUD, D., LE BACCON, P., OTTE, A. P., DISTECHE, C. M., AVNER, P. & HEARD, E. 2005. Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature,* 438**,** 369-73.

OKAMOTO, I., OTTE, A. P., ALLIS, C. D., REINBERG, D. & HEARD, E. 2004. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science,* 303**,** 644-9.

OKAMOTO, I., PATRAT, C., THEPOT, D., PEYNOT, N., FAUQUE, P., DANIEL, N., DIABANGOUAYA, P., WOLF, J. P., RENARD, J. P., DURANTHON, V. & HEARD, E. 2011. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature,* 472**,** 370-4.

PAGE, D. C., MOSHER, R., SIMPSON, E. M., FISHER, E. M. C., MARDON, G., POLLACK, J., MCGILLIVRAY, B., DE LA CHAPELLE, A. & BROWN, L. G. 1987. The sex-determining region of the human Y chromosome encodes a finger protein. *Cell,* 51**,** 1091-1104.

PALMER, M. S., SINCLAIR, A. H., BERTA, P., ELLIS, N. A., GOODFELLOW, P. N., ABBAS, N. E. & FELLOUS, M. 1989. Genetic evidence that ZFY is not the testis-determining factor. *Nature,* 342**,** 937-9.

PANNETIER, M., TILLY, G., KOCER, A., HUDRISIER, M., RENAULT, L., CHESNAIS, N., COSTA, J., LE PROVOST, F., VAIMAN, D., VILOTTE, J. L. & PAILHOUX, E. 2006. Goat SRY induces testis development in XX transgenic mice. *FEBS Lett,* 580**,** 3715-20.

PAQUET, D., KWART, D., CHEN, A., SPROUL, A., JACOB, S., TEO, S., OLSEN, K. M., GREGG, A., NOGGLE, S. & TESSIER-LAVIGNE, M. 2016. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature,* 533**,** 125-9.

PARK, S. H., SHIN, Y. K., SUH, Y. H., PARK, W. S., BAN, Y. L., CHOI, H. S., PARK, H. J. & JUNG, K. C. 2005. Rapid divergency of rodent CD99 orthologs: implications for the evolution of the pseudoautosomal region. *Gene,* 353**,** 177-88.

PENNY, G. D., KAY, G. F., SHEARDOWN, S. A., RASTAN, S. & BROCKDORFF, N. 1996. Requirement for Xist in X chromosome inactivation. *Nature,* 379**,** 131-7.

PESSIA, E., ENGELSTADTER, J. & MARAIS, G. A. 2014. The evolution of X chromosome inactivation in mammals: the demise of Ohno's hypothesis? *Cell Mol Life Sci,* 71**,** 1383-94.

PESSIA, E., MAKINO, T., BAILLY-BECHET, M., MCLYSAGHT, A. & MARAIS, G. A. 2012. Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proc Natl Acad Sci U S A,* 109**,** 5346-51.

PIERCE, A. J., HU, P., HAN, M., ELLIS, N. & JASIN, M. 2001. Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells. *Genes Dev,* 15**,** 3237-42.

PINDER, J., SALSMAN, J. & DELLAIRE, G. 2015. Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res,* 43**,** 9379-92.

PINKEL, D., GLEDHILL, B. L., LAKE, S., STEPHENSON, D. & VAN DILLA, M. A. 1982. Sex preselection in mammals? Separation of sperm bearing Y and "O" chromosomes in the vole Microtus oregoni. *Science,* 218**,** 904-6.

PLATT, R. J., CHEN, S., ZHOU, Y., YIM, M. J., SWIECH, L., KEMPTON, H. R., DAHLMAN, J. E., PARNAS, O., EISENHAURE, T. M., JOVANOVIC, M., GRAHAM, D. B., JHUNJHUNWALA, S., HEIDENREICH, M., XAVIER, R. J., LANGER, R., ANDERSON, D. G., HACOHEN, N., REGEV, A., FENG, G., SHARP, P. A. & ZHANG, F. 2014. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell,* 159**,** 440-55.

POMMIER, Y., LEO, E., ZHANG, H. & MARCHAND, C. 2010. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem Biol,* 17**,** 421-33.

PONTIER, D. B. & GRIBNAU, J. 2011. Xist regulation and function explored. *Hum Genet,* 130**,** 223-36.

POURCEL, C., SALVIGNOL, G. & VERGNAUD, G. 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology,* 151**,** 653-63.

PRITCHETT-CORNING, K. R., CLIFFORD, C. B. & FESTING, M. F. 2013. The effects of shipping on early pregnancy in laboratory rats. *Birth Defects Res B Dev Reprod Toxicol,* 98**,** 200-5.

QUADROS, R. M., MIURA, H., HARMS, D. W., AKATSUKA, H., SATO, T., AIDA, T., REDDER, R., RICHARDSON, G. P., INAGAKI, Y., SAKAI, D., BUCKLEY, S. M., SESHACHARYULU, P., BATRA, S. K., BEHLKE, M. A., ZEINER, S. A., JACOBI, A. M., IZU, Y., THORESON, W. B., URNESS, L. D., MANSOUR, S. L., OHTSUKA, M. & GURUMURTHY, C. B. 2017. Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. *Genome Biol,* 18**,** 92.

RALSTON, A. & ROSSANT, J. 2008. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Dev Biol,* 313**,** 614-29.

RAN, F. A., HSU, P. D., LIN, C. Y., GOOTENBERG, J. S., KONERMANN, S., TREVINO, A. E., SCOTT, D. A., INOUE, A., MATOBA, S., ZHANG, Y. & ZHANG, F. 2013a. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell,* 154**,** 1380-9.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013b. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc,* 8**,** 2281-308.

RASTAN, S. & ROBERTSON, E. J. 1985. X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. *J Embryol Exp Morphol,* 90**,** 379-88.

RATH, D., JOHNSON, L. A., DOBRINSKY, J. R., WELCH, G. R. & NIEMANN, H. 1997. Production of piglets preselected for sex following in vitro fertilization with

X and Y chromosome-bearing spermatozoa sorted by flow cytometry. *Theriogenology,* 47**,** 795-800.

RENFREE, M. B. & FENELON, J. C. 2017. The enigma of embryonic diapause. *Development,* 144**,** 3199-3210.

RENS, W., WELCH, G. R. & JOHNSON, L. A. 1998. A novel nozzle for more efficient sperm orientation to improve sorting efficiency of X and Y chromosome-bearing sperm. *Cytometry,* 33**,** 476-81.

RICE, W. R. 1984. Sex Chromosomes and the Evolution of Sexual Dimorphism. *Evolution,* 38**,** 735-742.

RICHARDSON, C. D., RAY, G. J., DEWITT, M. A., CURIE, G. L. & CORN, J. E. 2016. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol,* 34**,** 339-44.

ROPERS, H. H. 2010. Genetics of early onset cognitive impairment. *Annu Rev Genomics Hum Genet,* 11**,** 161-87.

ROSENBRUCH, M. 1994. [Early stages of the incubated chicken egg as a model in experimental biology and medicine]. *ALTEX,* 11**,** 199-206.

ROSENBRUCH, M. 1997. [The sensitivity of chicken embryos in incubated eggs]. *ALTEX,* 14**,** 111-113.

ROSS, M. T., GRAFHAM, D. V., COFFEY, A. J., SCHERER, S., MCLAY, K., MUZNY, D., PLATZER, M., HOWELL, G. R., BURROWS, C., BIRD, C. P., FRANKISH, A., LOVELL, F. L., HOWE, K. L., ASHURST, J. L., FULTON, R. S., SUDBRAK, R., WEN, G., JONES, M. C., HURLES, M. E., ANDREWS, T. D., SCOTT, C. E., SEARLE, S., RAMSER, J., WHITTAKER, A., DEADMAN, R., CARTER, N. P., HUNT, S. E., CHEN, R., CREE, A., GUNARATNE, P., HAVLAK, P., HODGSON, A., METZKER, M. L., RICHARDS, S., SCOTT, G., STEFFEN, D., SODERGREN, E., WHEELER, D. A., WORLEY, K. C., AINSCOUGH, R., AMBROSE, K. D., ANSARI-LARI, M. A., ARADHYA, S., ASHWELL, R. I., BABBAGE, A. K., BAGGULEY, C. L., BALLABIO, A., BANERJEE, R., BARKER, G. E., BARLOW, K. F., BARRETT, I. P., BATES, K. N., BEARE, D. M., BEASLEY, H., BEASLEY, O., BECK, A., BETHEL, G., BLECHSCHMIDT, K., BRADY, N., BRAY-ALLEN, S., BRIDGEMAN, A. M., BROWN, A. J., BROWN, M. J., BONNIN, D., BRUFORD, E. A., BUHAY, C., BURCH, P., BURFORD, D., BURGESS, J., BURRILL, W., BURTON, J., BYE, J. M., CARDER, C., CARREL, L., CHAKO, J., CHAPMAN, J. C., CHAVEZ, D., CHEN, E., CHEN, G., CHEN, Y., CHEN, Z., CHINAULT, C., CICCODICOLA, A., CLARK, S. Y., CLARKE, G., CLEE, C. M., CLEGG, S., CLERC-BLANKENBURG, K., CLIFFORD, K., COBLEY, V., COLE, C. G., CONQUER, J. S., CORBY, N., CONNOR, R. E., DAVID, R., DAVIES, J., DAVIS, C., DAVIS, J., DELGADO, O., DESHAZO, D., et al. 2005. The DNA sequence of the human X chromosome. *Nature,* 434**,** 325-37.

ROSSANT, J., NUTTER, L. M. & GERTSENSTEIN, M. 2011. Engineering the embryo. *Proc Natl Acad Sci U S A,* 108**,** 7659-60.

ROSSANT, J. & TAM, P. P. 2009. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development,* 136**,** 701-13.

ROUET, P., SMIH, F. & JASIN, M. 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol,* 14**,** 8096-106.

ROUGEULLE, C., CHAUMEIL, J., SARMA, K., ALLIS, C. D., REINBERG, D., AVNER, P. & HEARD, E. 2004. Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. *Mol Cell Biol,* 24**,** 5475-84.

ROZEN, S., SKALETSKY, H., MARSZALEK, J. D., MINX, P. J., CORDUM, H. S., WATERSTON, R. H., WILSON, R. K. & PAGE, D. C. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature,* 423**,** 873-6.

RULTEN, S. L., FISHER, A. E., ROBERT, I., ZUMA, M. C., ROULEAU, M., JU, L., POIRIER, G., REINA-SAN-MARTIN, B. & CALDECOTT, K. W. 2011. PARP-3 and APLF function together to accelerate nonhomologous end-joining. *Mol Cell,* 41**,** 33-45.

RUSSELL, L. B. & BANGHAM, J. W. 1961. Variegated-type position effects in the mouse. *Genetics,* 46**,** 509-25.

SADO, T., HOKI, Y. & SASAKI, H. 2006. Tsix defective in splicing is competent to establish Xist silencing. *Development,* 133**,** 4925-31.

SAMPSON, T. R. & WEISS, D. S. 2013. Cas9-dependent endogenous gene regulation is required for bacterial virulence. *Biochem Soc Trans,* 41**,** 1407-11.

SANGRITHI, M. N., ROYO, H., MAHADEVAIAH, S. K., OJARIKRE, O., BHAW, L., SESAY, A., PETERS, A. H., STADLER, M. & TURNER, J. M. 2017. Non-Canonical and Sexually Dimorphic X Dosage Compensation States in the Mouse and Human Germline. *Dev Cell,* 40**,** 289-301 e3.

SANGRITHI, M. N. & TURNER, J. M. A. 2018. Mammalian X Chromosome Dosage Compensation: Perspectives From the Germ Line. *Bioessays,* 40**,** e1800024.

SARTORI, A. A., LUKAS, C., COATES, J., MISTRIK, M., FU, S., BARTEK, J., BAER, R., LUKAS, J. & JACKSON, S. P. 2007. Human CtIP promotes DNA end resection. *Nature,* 450**,** 509-14.

SCHEININ, I., SIE, D., BENGTSSON, H., VAN DE WIEL, M. A., OLSHEN, A. B., VAN THUIJL, H. F., VAN ESSEN, H. F., EIJK, P. P., RUSTENBURG, F., MEIJER, G. A., REIJNEVELD, J. C., WESSELING, P., PINKEL, D., ALBERTSON, D. G. & YLSTRA, B. 2014. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res,* 24**,** 2022-32.

SCHULTZ, M. D., HE, Y., WHITAKER, J. W., HARIHARAN, M., MUKAMEL, E. A., LEUNG, D., RAJAGOPAL, N., NERY, J. R., URICH, M. A., CHEN, H., LIN, S., LIN, Y., JUNG, I., SCHMITT, A. D., SELVARAJ, S., REN, B., SEJNOWSKI, T. J., WANG, W. & ECKER, J. R. 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature,* 523**,** 212-6.

SEIDEL, G. E., JR. 2007. Overview of sexing sperm. *Theriogenology,* 68**,** 443-6.

SEIDEL, G. E., JR. 2012. Sexing mammalian sperm - Where do we go from here? *J Reprod Dev,* 58**,** 505-9.

SEKIDO, R. & LOVELL-BADGE, R. 2008. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature,* 453**,** 930-4.

SEREBROVSKY A, S. 1969. On the possibility of a new method for the control of insect pests. Vienna, Int. Atom. Energy Ag.

SHARMAN, G. B. 1971. Late DNA replication in the paternally derived X chromosome of female kangaroos. *Nature,* 230**,** 231-2.

SHARP, A. J., SPOTSWOOD, H. T., ROBINSON, D. O., TURNER, B. M. & JACOBS, P. A. 2002. Molecular and cytogenetic analysis of the spreading of X inactivation in X;autosome translocations. *Hum Mol Genet,* 11**,** 3145-56.

SHEN, B., ZHANG, J., WU, H., WANG, J., MA, K., LI, Z., ZHANG, X., ZHANG, P. & HUANG, X. 2013. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res,* 23**,** 720-3.

SHEN, M. W., ARBAB, M., HSU, J. Y., WORSTELL, D., CULBERTSON, S. J., KRABBE, O., CASSA, C. A., LIU, D. R., GIFFORD, D. K. & SHERWOOD, R. I. 2018. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature,* 563**,** 646-651.

SHIBATA, S. & LEE, J. T. 2004. Tsix transcription- versus RNA-based mechanisms in Xist repression and epigenetic choice. *Curr Biol,* 14**,** 1747-54.

SHIN, H. Y., WANG, C., LEE, H. K., YOO, K. H., ZENG, X., KUHNS, T., YANG, C. M., MOHR, T., LIU, C. & HENNIGHAUSEN, L. 2017. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat Commun,* 8**,** 15464.

SHPARGEL, K. B., SENGOKU, T., YOKOYAMA, S. & MAGNUSON, T. 2012. UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. *PLoS Genet,* 8**,** e1002964.

SIMON, M. D., PINTER, S. F., FANG, R., SARMA, K., RUTENBERG-SCHOENBERG, M., BOWMAN, S. K., KESNER, B. A., MAIER, V. K., KINGSTON, R. E. & LEE, J. T. 2013. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature,* 504**,** 465-469.

SIMPSON, A. J., CABALLERO, O. L., JUNGBLUTH, A., CHEN, Y. T. & OLD, L. J. 2005. Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer,* 5**,** 615-25.

SINCLAIR, A. H., BERTA, P., PALMER, M. S., HAWKINS, J. R., GRIFFITHS, B. L., SMITH, M. J., FOSTER, J. W., FRISCHAUF, A. M., LOVELL-BADGE, R. & GOODFELLOW, P. N. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature,* 346**,** 240-4.

SKALETSKY, H., KURODA-KAWAGUCHI, T., MINX, P. J., CORDUM, H. S., HILLIER, L., BROWN, L. G., REPPING, S., PYNTIKOVA, T., ALI, J., BIERI, T., CHINWALLA, A., DELEHAUNTY, A., DELEHAUNTY, K., DU, H., FEWELL, G., FULTON, L., FULTON, R., GRAVES, T., HOU, S. F., LATRIELLE, P., LEONARD, S., MARDIS, E., MAUPIN, R., MCPHERSON, J., MINER, T., NASH, W., NGUYEN, C., OZERSKY, P., PEPIN, K., ROCK, S., ROHLFING, T., SCOTT, K., SCHULTZ, B., STRONG, C., TIN-WOLLAM, A., YANG, S. P., WATERSTON, R. H., WILSON, R. K., ROZEN, S. & PAGE, D. C. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature,* 423**,** 825-37.

SKUSE, D. H. 2005. X-linked genes and mental functioning. *Hum Mol Genet,* 14 Spec No 1**,** R27-32.

SNELL, D. M. & TURNER, J. M. A. 2018. Sex Chromosome Effects on Male-Female Differences in Mammals. *Curr Biol,* 28**,** R1313-R1324.

SOH, Y. Q., ALFOLDI, J., PYNTIKOVA, T., BROWN, L. G., GRAVES, T., MINX, P. J., FULTON, R. S., KREMITZKI, C., KOUTSEVA, N., MUELLER, J. L., ROZEN, S., HUGHES, J. F., OWENS, E., WOMACK, J. E., MURPHY, W. J.,

CAO, Q., DE JONG, P., WARREN, W. C., WILSON, R. K., SKALETSKY, H. & PAGE, D. C. 2014. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell,* 159**,** 800-13.

SONG, J., YANG, D., XU, J., ZHU, T., CHEN, Y. E. & ZHANG, J. 2016. RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat Commun,* 7**,** 10548.

SORIANO, P. 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet,* 21**,** 70-1.

SPEKSNIJDER, G. & IVARIE, R. 2000. A modified method of shell windowing for producing somatic or germline chimeras in fertilized chicken eggs. *Poult Sci,* 79**,** 1430-3.

SPIEGEL, A., BACHMANN, M., JURADO JIMENEZ, G. & SAROV, M. 2019. CRISPR/Cas9-based knockout pipeline for reverse genetics in mammalian cell culture. *Methods,* 164-165**,** 49-58.

SRINIVAS, S., WATANABE, T., LIN, C. S., WILLIAM, C. M., TANABE, Y., JESSELL, T. M. & COSTANTINI, F. 2001. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev Biol,* 1**,** 4.

SRIVASTAVA, M., NAMBIAR, M., SHARMA, S., KARKI, S. S., GOLDSMITH, G., HEGDE, M., KUMAR, S., PANDEY, M., SINGH, R. K., RAY, P., NATARAJAN, R., KELKAR, M., DE, A., CHOUDHARY, B. & RAGHAVAN, S. C. 2012. An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell,* 151**,** 1474-87.

STAVROPOULOS, N., LU, N. & LEE, J. T. 2001. A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proc Natl Acad Sci U S A,* 98**,** 10232-7.

STERNBERG, N. & HAMILTON, D. 1981. Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. *J Mol Biol,* 150**,** 467-86.

STEVENS, N. M. 1905a. *Studies in spermatogenesis : with especial reference to the "accessory chromosome",* Washington, D.C., Carnegie Institution.

STEVENS, N. M. 1905b. A study of the germ cells of Aphis rosæ and Aphis œnotheræ. *Journal of Experimental Zoology,* 2**,** 313-333.

STEVENSON, B. J., ISELI, C., PANJI, S., ZAHN-ZABAL, M., HIDE, W., OLD, L. J., SIMPSON, A. J. & JONGENEEL, C. V. 2007. Rapid evolution of cancer/testis genes on the X chromosome. *BMC Genomics,* 8**,** 129.

STRUMPF, D., MAO, C. A., YAMANAKA, Y., RALSTON, A., CHAWENGSAKSOPHAK, K., BECK, F. & ROSSANT, J. 2005. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development,* 132**,** 2093-102.

SYMINGTON, L. S. & GAUTIER, J. 2011. Double-strand break end resection and repair pathway choice. *Annu Rev Genet,* 45**,** 247-71.

SZOSTAK, J. W., ORR-WEAVER, T. L., ROTHSTEIN, R. J. & STAHL, F. W. 1983. The double-strand-break repair model for recombination. *Cell,* 33**,** 25-35.

TADA, T., OBATA, Y., TADA, M., GOTO, Y., NAKATSUJI, N., TAN, S., KONO, T. & TAKAGI, N. 2000. Imprint switching for non-random X-chromosome inactivation during mouse oocyte growth. *Development,* 127**,** 3101-5.

TAKAGI, N. & SASAKI, M. 1975. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature,* 256**,** 640-2.

TANG, S. H., SILVA, F. J., TSARK, W. M. & MANN, J. R. 2002. A Cre/loxP-deleter transgenic line in mouse strain 129S1/SvImJ. *Genesis,* 32**,** 199-202.

TASIC, B., HIPPENMEYER, S., WANG, C., GAMBOA, M., ZONG, H., CHEN-TSAI, Y. & LUO, L. 2011. Site-specific integrase-mediated transgenesis in mice via pronuclear injection. *Proc Natl Acad Sci U S A,* 108**,** 7902-7.

TIAN, D., SUN, S. & LEE, J. T. 2010. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell,* 143**,** 390-403.

TODER, R. & GRAVES, J. A. 1998. CSF2RA, ANT3, and STS are autosomal in marsupials: implications for the origin of the pseudoautosomal region of mammalian sex chromosomes. *Mamm Genome,* 9**,** 373-6.

TORRES, R. J. & PUIG, J. G. 2007. Hypoxanthine-guanine phosophoribosyltransferase (HPRT) deficiency: Lesch-Nyhan syndrome. *Orphanet journal of rare diseases,* 2**,** 48-48.

TRUJILLO, K. M., YUAN, S. S., LEE, E. Y. & SUNG, P. 1998. Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. *J Biol Chem,* 273**,** 21447-50.

TUBMAN, L. M., BRINK, Z., SUH, T. K. & SEIDEL, G. E., JR. 2004. Characteristics of calves produced with sperm sexed by flow cytometry/cell sorting. *J Anim Sci,* 82**,** 1029-36.

TUDURI, S., CRABBÉ, L., CONTI, C., TOURRIÈRE, H., HOLTGREVE-GREZ, H., JAUCH, A., PANTESCO, V., DE VOS, J., THOMAS, A., THEILLET, C., POMMIER, Y., TAZI, J., COQUELLE, A. & PASERO, P. 2009. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nature cell biology,* 11**,** 1315-1324.

TUKIAINEN, T., VILLANI, A.-C., YEN, A., RIVAS, M. A., MARSHALL, J. L., SATIJA, R., AGUIRRE, M., GAUTHIER, L., FLEHARTY, M., KIRBY, A., CUMMINGS, B. B., CASTEL, S. E., KARCZEWSKI, K. J., AGUET, F., BYRNES, A., CONSORTIUM, G. T., AGUET, F., ARDLIE, K. G., CUMMINGS, B. B., GELFAND, E. T., GETZ, G., HADLEY, K., HANDSAKER, R. E., HUANG, K. H., KASHIN, S., KARCZEWSKI, K. J., LEK, M., LI, X., MACARTHUR, D. G., NEDZEL, J. L., NGUYEN, D. T., NOBLE, M. S., SEGRÈ, A. V., TROWBRIDGE, C. A., TUKIAINEN, T., ABELL, N. S., BALLIU, B., BARSHIR, R., BASHA, O., BATTLE, A., BOGU, G. K., BROWN, A., BROWN, C. D., CASTEL, S. E., CHEN, L. S., CHIANG, C., CONRAD, D. F., COX, N. J., DAMANI, F. N., DAVIS, J. R., DELANEAU, O., DERMITZAKIS, E. T., ENGELHARDT, B. E., ESKIN, E., FERREIRA, P. G., FRÉSARD, L., GAMAZON, E. R., GARRIDO-MARTÍN, D., GEWIRTZ, A. D. H., GLINER, G., GLOUDEMANS, M. J., GUIGO, R., HALL, I. M., HAN, B., HE, Y., HORMOZDIARI, F., HOWALD, C., KYUNG IM, H., JO, B., YONG KANG, E., KIM, Y., KIM-HELLMUTH, S., LAPPALAINEN, T., LI, G., LI, X., LIU, B., MANGUL, S., MCCARTHY, M. I., MCDOWELL, I. C., MOHAMMADI, P., MONLONG, J., MONTGOMERY, S. B., MUÑOZ-AGUIRRE, M., NDUNGU, A. W., NICOLAE, D. L., NOBEL, A. B., OLIVA, M., ONGEN, H., PALOWITCH, J. J., PANOUSIS, N., PAPASAIKAS, P., PARK, Y., PARSANA, P., PAYNE, A. J., PETERSON, C. B., QUAN, J.,

REVERTER, F., SABATTI, C., SAHA, A., SAMMETH, M., et al. 2017. Landscape of X chromosome inactivation across human tissues. *Nature,* 550**,** 244.

TURNER, J. M. 2007. Meiotic sex chromosome inactivation. *Development,* 134**,** 1823-31.

TURNER, J. M., MAHADEVAIAH, S. K., ELLIOTT, D. J., GARCHON, H. J., PEHRSON, J. R., JAENISCH, R. & BURGOYNE, P. S. 2002. Meiotic sex chromosome inactivation in male mice with targeted disruptions of Xist. *J Cell Sci,* 115**,** 4097-105.

TURNER, J. M., MAHADEVAIAH, S. K., ELLIS, P. J., MITCHELL, M. J. & BURGOYNE, P. S. 2006. Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial postmeiotic repression in spermatids. *Dev Cell,* 10**,** 521-9.

UMEHARA, T., TSUJITA, N. & SHIMADA, M. 2019. Activation of Toll-like receptor 7/8 encoded by the X chromosome alters sperm motility and provides a novel simple technology for sexing sperm. *PLoS Biol,* 17**,** e3000398.

UNCKLESS, R. L., CLARK, A. G. & MESSER, P. W. 2017. Evolution of Resistance Against CRISPR/Cas9 Gene Drive. *Genetics,* 205**,** 827-841.

UNTERGASSER, A., CUTCUTACHE, I., KORESSAAR, T., YE, J., FAIRCLOTH, B. C., REMM, M. & ROZEN, S. G. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res,* 40**,** e115.

USUI, T., OHTA, T., OSHIUMI, H., TOMIZAWA, J., OGAWA, H. & OGAWA, T. 1998. Complex formation and functional versatility of Mre11 of budding yeast in recombination. *Cell,* 95**,** 705-16.

VAN OVERBEEK, M., CAPURSO, D., CARTER, M. M., THOMPSON, M. S., FRIAS, E., RUSS, C., REECE-HOYES, J. S., NYE, C., GRADIA, S., VIDAL, B., ZHENG, J., HOFFMAN, G. R., FULLER, C. K. & MAY, A. P. 2016. DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol Cell,* 63**,** 633-646.

VEYRUNES, F., WATERS, P. D., MIETHKE, P., RENS, W., MCMILLAN, D., ALSOP, A. E., GRUTZNER, F., DEAKIN, J. E., WHITTINGTON, C. M., SCHATZKAMER, K., KREMITZKI, C. L., GRAVES, T., FERGUSON-SMITH, M. A., WARREN, W. & MARSHALL GRAVES, J. A. 2008. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res,* 18**,** 965-73.

VICOSO, B. & CHARLESWORTH, B. 2006. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet,* 7**,** 645-53.

VICOSO, B. & CHARLESWORTH, B. 2009. EFFECTIVE POPULATION SIZE AND THE FASTER-X EFFECT: AN EXTENDED MODEL. *Evolution,* 63**,** 2413-2426.

VIDAL, V. P., CHABOISSIER, M. C., DE ROOIJ, D. G. & SCHEDL, A. 2001. Sox9 induces testis development in XX transgenic mice. *Nat Genet,* 28**,** 216-7.

WALKER, J. R., CORPINA, R. A. & GOLDBERG, J. 2001. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature,* 412**,** 607-14.

WALPORT, L. J., HOPKINSON, R. J., VOLLMAR, M., MADDEN, S. K., GILEADI, C., OPPERMANN, U., SCHOFIELD, C. J. & JOHANSSON, C. 2014. Human UTY(KDM6C) is a male-specific N-methyl lysyl demethylase. *J Biol Chem,* 289**,** 18302-13.

WANG, H., HU, Y. C., MARKOULAKI, S., WELSTEAD, G. G., CHENG, A. W., SHIVALILA, C. S., PYNTIKOVA, T., DADON, D. B., VOYTAS, D. F., BOGDANOVE, A. J., PAGE, D. C. & JAENISCH, R. 2013a. TALEN-mediated editing of the mouse Y chromosome. *Nat Biotechnol,* 31**,** 530-2.

WANG, H., YANG, H., SHIVALILA, C. S., DAWLATY, M. M., CHENG, A. W., ZHANG, F. & JAENISCH, R. 2013b. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell,* 153**,** 910-8.

WANG, J. C. 2002. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol,* 3**,** 430-40.

WANG, M., SUN, Z., ZOU, Z., DING, F., LI, L., WANG, H., ZHAO, C., LI, N. & DAI, Y. 2018. Efficient targeted integration into the bovine Rosa26 locus using TALENs. *Scientific Reports,* 8**,** 10385.

WANG, P. J., MCCARREY, J. R., YANG, F. & PAGE, D. C. 2001. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet,* 27**,** 422-6.

WASSARMAN, P. M. & KINLOCH, R. A. 1992. Gene expression during oogenesis in mice. *Mutat Res,* 296**,** 3-15.

WATERS, P. D., WALLIS, M. C. & MARSHALL GRAVES, J. A. 2007. Mammalian sex--Origin and evolution of the Y chromosome and SRY. *Semin Cell Dev Biol,* 18**,** 389-400.

WEBSTER, B., HAYES, W. & PIKE, T. W. 2015. Avian egg odour encodes information on embryo sex, fertility and development. *PLoS One,* 10**,** e0116345.

WEICHERT, C. K. 1940. The experimental shortening of delayed pregnancy in the albino rat. *Anatomical Record,* 77**,** 31-47.

WEICHERT, C. K. 1942. The experimental control of prolonged pregnancy in the lactating rat by means of estrogen. *The Anatomical Record,* 83**,** 1-17.

WEISSMANN, A., REITEMEIER, S., HAHN, A., GOTTSCHALK, J. & EINSPANIER, A. 2013. Sexing domestic chicken before hatch: a new method for in ovo gender identification. *Theriogenology,* 80**,** 199-205.

WEST, J. D., FRELS, W. I., CHAPMAN, V. M. & PAPAIOANNOU, V. E. 1977. Preferential expression of the maternally derived X chromosome in the mouse yolk sac. *Cell,* 12**,** 873-82.

WETERINGS, E. & VAN GENT, D. C. 2004. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair (Amst),* 3**,** 1425-35.

WILLISON, K. R. & LYON, M. F. 2000. A UK-centric history of studies on the mouse t-complex. *Int J Dev Biol,* 44**,** 57-63.

WILSON, E. B. 1909. Recent Researches on the Determination and Heredity of Sex. *Science,* 29**,** 53-70.

WINDBICHLER, N., MENICHELLI, M., PAPATHANOS, P. A., THYME, S. B., LI, H., ULGE, U. Y., HOVDE, B. T., BAKER, D., MONNAT, R. J., JR., BURT, A. & CRISANTI, A. 2011. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature,* 473**,** 212-5.

WINDBICHLER, N., PAPATHANOS, P. A., CATTERUCCIA, F., RANSON, H., BURT, A. & CRISANTI, A. 2007. Homing endonuclease mediated gene targeting in Anopheles gambiae cells and embryos. *Nucleic Acids Res,* 35**,** 5922-33.

WINDBICHLER, N., PAPATHANOS, P. A. & CRISANTI, A. 2008. Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio

distortion and early dominant embryo lethality in Anopheles gambiae. *PLoS Genet,* 4**,** e1000291.

WRIGHT, C. M., VAN DER MERWE, M., DEBROT, A. H. & BJORNSTI, M. A. 2015. DNA topoisomerase I domain interactions impact enzyme activity and sensitivity to camptothecin. *J Biol Chem,* 290**,** 12068-78.

WUTZ, A. & JAENISCH, R. 2000. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol Cell,* 5**,** 695-705.

XIONG, Y., CHEN, X., CHEN, Z., WANG, X., SHI, S., WANG, X., ZHANG, J. & HE, X. 2010. RNA sequencing shows no dosage compensation of the active X-chromosome. *Nat Genet,* 42**,** 1043-7.

YAMAUCHI, Y., RIEL, J. M., STOYTCHEVA, Z. & WARD, M. A. 2014. Two Y genes can replace the entire Y chromosome for assisted reproduction in the mouse. *Science,* 343**,** 69-72.

YANG, D., SONG, J., ZHANG, J., XU, J., ZHU, T., WANG, Z., LAI, L. & CHEN, Y. E. 2016. Identification and characterization of rabbit ROSA26 for gene knock-in and stable reporter gene expression. *Scientific reports,* 6**,** 25161-25161.

YANG, F., BABAK, T., SHENDURE, J. & DISTECHE, C. M. 2010. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res,* 20**,** 614-22.

YANG, H., WANG, H., SHIVALILA, C. S., CHENG, A. W., SHI, L. & JAENISCH, R. 2013a. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell,* 154**,** 1370-9.

YANG, L., GUELL, M., BYRNE, S., YANG, J. L., DE LOS ANGELES, A., MALI, P., AACH, J., KIM-KISELAK, C., BRIGGS, A. W., RIOS, X., HUANG, P. Y., DALEY, G. & CHURCH, G. 2013b. Optimization of scarless human stem cell genome editing. *Nucleic Acids Res,* 41**,** 9049-61.

YANG, X., LENG, X., TU, W., LIU, Y., XU, J., PEI, X., MA, Y., YANG, D. & YANG, Y. 2018. Spermatogenic phenotype of testis-specific protein, Y-encoded, 1 (TSPY1) dosage deficiency is independent of variations in TSPY-like 1 (TSPYL1) and TSPY-like 5 (TSPYL5): a case-control study in a Han Chinese population. *Reprod Fertil Dev,* 30**,** 555-562.

YILDIRIM, E., SADREYEV, R. I., PINTER, S. F. & LEE, J. T. 2011. X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription. *Nature Structural &Amp; Molecular Biology,* 19**,** 56.

YİLMAZ-DİKMEN, B. & DİKMEN, S. 2013. A morphometric method of sexing white layer eggs. *Brazilian Journal of Poultry Science,* 15**,** 203-210.

YIN, M., JIANG, W., FANG, Z., KONG, P., XING, F., LI, Y., CHEN, X. & LI, S. 2015. Generation of hypoxanthine phosphoribosyltransferase gene knockout rabbits by homologous recombination and gene trapping through somatic cell nuclear transfer. *Scientific Reports,* 5**,** 16023.

YOSEF, I., EDRY-BOTZER, L., GLOBUS, R., SHLOMOVITZ, I., MUNITZ, A., GERLIC, M. & QIMRON, U. 2019. A genetic system for biasing the sex ratio in mice. *EMBO Rep***,** e48269.

YOSHIMI, K., KUNIHIRO, Y., KANEKO, T., NAGAHORA, H., VOIGT, B. & MASHIMO, T. 2016. ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nat Commun,* 7**,** 10431.

ZECHNER, U., WILDA, M., KEHRER-SAWATZKI, H., VOGEL, W., FUNDELE, R. & HAMEISTER, H. 2001. A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution? *Trends Genet,* 17**,** 697-701.

ZHANG, Z., NIU, B., JI, D., LI, M., LI, K., JAMES, A. A., TAN, A. & HUANG, Y. 2018. Silkworm genetic sexing through W chromosome-linked, targeted gene integration. *Proc Natl Acad Sci U S A,* 115**,** 8752-8756.

ZHAO, X., WEI, C., LI, J., XING, P., LI, J., ZHENG, S. & CHEN, X. 2017. Cell cycle-dependent control of homologous recombination. *Acta Biochim Biophys Sin (Shanghai),* 49**,** 655-668.

ZHU, F., GAMBOA, M., FARRUGGIO, A. P., HIPPENMEYER, S., TASIC, B., SCHULE, B., CHEN-TSAI, Y. & CALOS, M. P. 2014. DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. *Nucleic Acids Res,* 42**,** e34.

ZHU, Z., CHUNG, W. H., SHIM, E. Y., LEE, S. E. & IRA, G. 2008. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell,* 134**,** 981-94.

ZUCCOTTI, M., BOIANI, M., PONCE, R., GUIZZARDI, S., SCANDROGLIO, R., GARAGNA, S. & REDI, C. A. 2002. Mouse Xist expression begins at zygotic genome activation and is timed by a zygotic clock. *Mol Reprod Dev,* 61**,** 14-20.

ZVETKOVA, I., APEDAILE, A., RAMSAHOYE, B., MERMOUD, J. E., CROMPTON, L. A., JOHN, R., FEIL, R. & BROCKDORFF, N. 2005. Global hypomethylation of the genome in XX embryonic stem cells. *Nature Genetics,* 37**,** 1274-1279.